

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 8, 2002, 12:27:08 ; Search time 1619 Seconds
(without alignments)
2300.902 Million cell updates/sec

Title: US-09-065-672-5_COPY_1_276
Perfect score: 276
Sequence: 1 CTAGCGCGTCAACAGAGC.....CGGAGCGCGAGCAGAGC 276

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 6748477542 residues

Word size : 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

EST: *
1: em_estb: *
2: em_estb: *
3: em_estb: *
4: em_estb: *
5: em_estb: *
6: em_estb: *
7: em_estb: *
8: em_estb: *
9: gb_est: *
10: gb_est: *
11: gb_est: *
12: gb_est: *
13: em_gss_hum: *
14: em_gss_inv: *
15: em_gss_pln: *
16: em_gss_vrt: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	225	81.5	1074	10	BG611169 602612141
2	223	80.8	369	9	AT792251 np78b10.y
3	222	80.4	329	9	AT682287 wc51c01.x
4	222	80.4	375	9	AA578209 n156f11.s
5	222	80.4	380	9	AA631916 np78b10.s
6	222	80.4	642	9	AT557474 PT2.1.7-G
7	191	69.2	294	9	AT417931 tg55e07.x
8	189	68.5	760	9	AT557019 PT2.1.10-
9	187	67.8	378	9	AA876897 ny47g12.s
10	174	63.0	742	10	BG612025 602613927
11	165	59.8	297	10	BF438241 7g01c02.x
12	131	47.5	322	9	AA618586 np30h03.s
13	84	30.4	551	10	BG498699 602544315
14	76	27.5	370	9	AT972706 wr42d04.x
15	53	19.2	545	12	AA503931 RPECI-11-3
16	22	8.0	249	9	AA330052 EST33753
17	22	8.0	1010	10	BF981605 602305903

C 18	21	7.6	281	10	BF941269	BF941269 7d96b10.x
C 19	21	7.6	396	12	AQ470489	AQ470489 CTIB1-E1-
C 20	21	7.6	594	9	AM817087	AM817087 OVO-ST024
C 21	20	7.2	189	10	BG977650	BG977650 MR2-CT018
C 22	20	7.2	232	9	AM847033	AM847033 RCI-CT019
C 23	20	7.2	245	10	BG988262	BG988262 PMO-HT116
C 24	20	7.2	292	9	AA076979	AA076979 7B05G09 C
C 25	20	7.2	298	9	AA369232	AA369232 EST80614
C 26	20	7.2	304	9	AM904887	AM904887 RC5-NN106
C 27	20	7.2	333	12	AQ097836	AQ097836 HS-3038_A
C 28	20	7.2	355	9	AT420666	AT420666 CT13C08.x
C 29	20	7.2	357	10	R84335	R84335 Y923005.r1
C 30	20	7.2	368	9	AA535756	AA535756 n178d01.s
C 31	20	7.2	370	10	T88871	T88871 yd31a10.s1
C 32	20	7.2	372	9	AT079575	AT079575 oz04h10.x
C 33	20	7.2	388	10	BF68361	BF68361 I15-ET011
C 34	20	7.2	388	12	AQ041496	AQ041496 CIT-HSP-2
C 35	20	7.2	407	9	AT473651	AT473651 tm02e11.x
C 36	20	7.2	417	10	BE767059	BE767059 RC4-NT011
C 37	20	7.2	419	9	AM243817	AM243817 xo57a08.x
C 38	20	7.2	427	10	N58756	N58756 yv75a10.s1
C 39	20	7.2	431	9	AA600127	AA600127 ae50b06.s
C 40	20	7.2	450	12	AQ338607	AQ338607 HS-2202_B
C 41	20	7.2	453	10	H05904	H05904 y171c02.s1
C 42	20	7.2	471	9	AT038577	AT038577 ox34c10.s
C 43	20	7.2	484	10	BF884356	BF884356 RC3-ET013
C 44	20	7.2	488	9	BE148220	BE148220 MR0-HT024
C 45	20	7.2	493	9	AA703950	AA703950 ag79e03.r

ALIGNMENTS

RESULT 1
BG611169 1074 bp mRNA linear EST 18-APR-2001
LOCUS 602612141F1 NIH_MGC_60 Homo sapiens cDNA clone IMAGE:4737322 5',
DEFINITION BG611169
ACCESSION BG611169
VERSION BG611169.1 GI:13662540
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE 1 (bases 1 to 1074)
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov

Tissue Procurement: DCTD/DRP
CDNA Library Preparation: CLONETECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
http://image.llnl.gov
Plate: LNCM1605 row: b column: 11
High quality sequence stop: 469.
Location/Qualifiers

FEATURES

1..1074
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="4737322"
/clone_lib="NIH_MGC_60"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: prostate; Vector: pDNR-LIB (Clontech);
Site:1: SfiI (ggcgctcgccg); Site:2: SfiI (ggcgctatggcc
); Double-stranded cDNA was prepared from cell line RNA.
5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor

sequence: 5'-ATTCTAGAGCCGAGGCGCCGACATC-dt(30)BN-3'
(where B = A, C, or G, and N = A, C, G, or T). Average
insert size 1.5 kb (range 0.9-4.0 kb). 14/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA). Note: this is a NIH-MGC
Library."

BASE COUNT 416 a 209 c 307 g 142 t
ORIGIN

Query Match 81.5%; Score 225; DB 10; Length 1074;
Best Local Similarity 99.6%; Pred. No. 1.5e-108;

Matches 275; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTAAGGCGTGAACAGAGAGCCGACCTGGAGGCTGAACCTTTAGGCCGATGCTCTTG 60
|||||
Db 21 CTAAGGCGTGAACAGAGAGCCGACCTGGAGGCTGAACCTTTAGGCCGATGCTCTTG 80
Qy 61 CAAGGTCAGGCAAGCTGGATTGCTGCCACCTTTGGCAGAGAGAACAGCAGATGTTGTC 120
|||||
Db 81 CAAGGTCAGGCAAGCTGGATTGCTGCCACCTTTGGCAGAGAGAACAGCAGATGTTGTC 140
Qy 121 GCCCATTTCTGATCAAGAGCCGCCCATCTTACTACCTCCAGAGTCTTTCTCTCT 180
|||||
Db 141 GCCCATTTCTGATCAAGAGCCGCCCATCTTACTACCTCCAGAGTCTTTCTCTCT 200
Qy 181 AATAAGAAACATCTACTTGAACATCTACTGGGCGAGACAGAGATGATGCTCAGCC 240
|||||
Db 201 AATAAGAAACATCTACTTGAACATCTACTGGGCGAGACAGAGATGATGCTCAGCC 260
Qy 241 TCTAATTCGTGAATTTGGGAGGCGGAGCAGAGAG 276
|||||
Db 261 TCTAATTCGTGAATTTGGGAGGCGGAGCAGAGAG 296

RESULT 2 369 bp mRNA linear EST 13-DEC-1999
A1792251
LOCUS np8B10.y5 NCI-CGAP.Pr2 Homo sapiens cDNA clone IMAGE:1132411

DEFINITION similar to contains Alu repetitive element; contains element MER4
repetitive element ;, mRNA sequence.

ACCESSION A1792251
VERSION A1792251.1 GI:5339967
KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Other ESTs: np78B10.x5

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuahqui, M.D.

, Michael Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: David B. Krizman, Ph.D.

cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

www-bio.lnl.gov/db/rrp/image/image.html

This read is a RESEQUENCE of a previously sequenced human clone
Original clone citation: see original entry for original citation
information
This 5' resequenced clone has no previous 5' data to verify this
new read against
Putative full length read
The vector to vector length is 370
Insert Length: 470 Std Error: 0.00

FEATURES Seq primer: -40RP from Gibco.
Location/Qualifiers
source 1..369

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1132411"
/clone_11b="NCI-CGAP.Pr2"
/sex="Male"
/dev_stage="45 years old"
/lab_host="DH10B"
/note="Vector: PAMP10; Site 1: NotI; Site 2: EcoRI; 1st
strand cDNA was primed with oligo(dT)₁₇ on 50 ng of
DNase-treated, total cellular RNA obtained from 5,000-10
000 microdissected preneoplastic cells
histologically-determined to be prostatic intraepithelial
neoplasia 2 (PIN2) cells. Double-stranded cDNA was
ligated to EcoRI adaptors, 5 cycles of PCR applied to the
cDNA with an adaptor-specific primer, and the resulting
PCR product subcloned into PAMP10 by the UDG-cloning
method (Life Technologies). Average insert size is 600
bp. NOTE: Not directionally cloned. This library was
constructed by David Krizman."

BASE COUNT 101 a 88 c 93 g 87 t
ORIGIN

Query Match 80.8%; Score 223; DB 9; Length 369;
Best Local Similarity 99.6%; Pred. No. 1.4e-107;

Matches 273; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 AAGGCGTGAACAGAGAGCCGACCTGGAGGCTGAACCTTTAGGCCGATGCTCTTGA 62
|||||
Db 7 AAGGCGTGAACAGAGAGCCGACCTGGAGGCTGAACCTTTAGGCCGATGCTCTTGA 66
Qy 63 AGGTCAGCAAGCTGATCTGTGCTCCACCTTTGCAGAGAGAACAGCATGTTGTGCC 122
|||||
Db 67 AGGTCAGCAAGCTGATCTGTGCTCCACCTTTGCAGAGAGAACAGCATGTTGTGCC 126
Qy 123 CCATTTCTGATCAAGAGAGCCGCCCATCTTACTACCTCCAGAGTCTTTCTCTTA 182
|||||
Db 127 CCATTTCTGATCAAGAGAGCCGCCCATCTTACTACCTCCAGAGTCTTTCTCTTA 186
Qy 183 TAAGAAACATCTACTTGAACATCTACTGGGCGAGACAGAGATGATGCTCAGCC 242
|||||
Db 187 TAAGAAACATCTACTTGAACATCTACTGGGCGAGACAGAGATGATGCTCAGCC 246
Qy 243 TAATTCGTGAATTTGGGAGGCGGAGCAGAGAG 276
|||||
Db 247 TAATTCGTGAATTTGGGAGGCGGAGCAGAGAG 280

RESULT 3 329 bp mRNA linear EST 17-DEC-1999
A1682287/c
LOCUS wc51c01.x1 NCI-CGAP.Pr28 Homo sapiens cDNA clone IMAGE:2322144 3'

DEFINITION similar to contains Alu repetitive element; contains element MER4
repetitive element ;, mRNA sequence.

ACCESSION A1682287
VERSION A1682287.1 GI:4892469
KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbdp/image/image.html

Insert Length: 412 Std Error: 0.00
Seq primer: -40UP from Gibco.

FEATURES

source

Location/Qualifiers

1..329

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:2322144"

/clone_1lb="NCI_CGAP_Pr28"

/sex="male"

/dev_stage="adult"

/lab_host="DH10B"

/note="Organ: prostate; Vector: pT7g3D-Pac (Pharmacia)

with a modified polylinker; Plasmid DNA from the

normalized library NCI-CGAP-Pr22 was prepared, and as

clones were made in vitro. Following HAP purification,

this DNA was used as tracer in a subtractive hybridization

reaction. The driver was PCR-amplified cDNAs from a pool

of 5,000 clones made from the same library (clones

985608-986759, 1101192-1101959, and 1217928-1220615).

Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT

83 a 78 c 78 g 90 t

Query Match

Best Local Similarity 100.0%; Score 222; DB 9; Length 329;

Matches 222; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 TGTGTGCAAGTCAGGCAAGCTGATTCGTGTCCTCCACCTTTGAGAGAGAACGCATG 114

Db 310 TGTGTGCAAGTCAGGCAAGCTGATTCGTGTCCTCCACCTTTGAGAGAGAACGCATG 251

QY 115 TTGTGCGCCATTTCAGATCAAGACCGCCCATCTTACTACCTCAAGAGTCTTTT 174

Db 250 TTGTGCGCCATTTCAGATCAAGACCGCCCATCTTACTACCTCAAGAGTCTTTT 191

QY 175 CTCTCTAATAGAAACATCTACTTTGAAACATCTACTGGGCGAGACGAGTGTATGC 234

Db 190 CTCTCTAATAGAAACATCTACTTTGAAACATCTACTGGGCGAGACGAGTGTATGC 131

QY 235 TCAGCCTGTAATTCGTGAATTCGGGAGCGCCGAGGAGAGAG 276

Db 130 TCAGCCTGTAATTCGTGAATTCGGGAGCGCCGAGGAGAGAG 89

RESULT 4
AA578209 375 bp mRNA linear EST 12-SEP-1997

LOCUS n13611.s1 NCI_CGAP-Pr4 Homo sapiens cDNA clone IMAGE:1044717

DEFINITION similar to contains Alu repetitive element; contains element MER4

repetitive element ;, mRNA sequence.

ACCESSION AA578209

VERSION AA578209.1 GI:2356393

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 375)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

Authors National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgaps-remail.nih.gov

Tissue Procurement: W. Marston Linehan, M.D., Rodrigo F. Chuagui,

M.D., Michael R. Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: David B. Kitzman, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbdp/image/image.html

Insert Length: 395 Std Error: 0.00
Seq primer: -40ml3 fwd. RT from Amersham.

FEATURES

source

Location/Qualifiers

1..375

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:1044717"

/clone_1lb="NCI_CGAP_Pr4"

/sex="male"

/tissue_type="prostatic intraepithelial neoplasia - high

grade"

/dev_stage="adult"

/lab_host="DH10B"

/note="Organ: prostate; Vector: PAMPI0; mRNA made from

prostate intraepithelial neoplasia (high-grade), cDNA

made by oligo-dT priming. Non-directionally cloned.

Size-selected on agarose gel, average insert size 600 bp.

BASE COUNT

110 a 87 c 90 g 88 t

Query Match

Best Local Similarity 100.0%; Score 222; DB 9; Length 375;

Matches 222; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 TGTGTGCAAGTCAGGCAAGCTGATTCGTGTCCTCCACCTTTGAGAGAGAACGCATG 114

Db 48 TGTGTGCAAGTCAGGCAAGCTGATTCGTGTCCTCCACCTTTGAGAGAGAACGCATG 107

QY 115 TTGTGCGCCATTTCAGATCAAGACCGCCCATCTTACTACCTCAAGAGTCTTTT 174

Db 108 TTGTGCGCCATTTCAGATCAAGACCGCCCATCTTACTACCTCAAGAGTCTTTT 167

QY 175 CTCTCTAATAGAAACATCTACTTTGAAACATCTACTGGGCGAGACGAGTGTATGC 234

Db 168 CTCTCTAATAGAAACATCTACTTTGAAACATCTACTGGGCGAGACGAGTGTATGC 227

QY 235 TCAGCCTGTAATTCGTGAATTCGGGAGCGCCGAGGAGAGAG 276

Db 228 TCAGCCTGTAATTCGTGAATTCGGGAGCGCCGAGGAGAGAG 269

RESULT 5
AA631916/c 380 bp mRNA linear EST 30-OCT-1997

LOCUS np78b10.s1 NCI_CGAP-Pr2 Homo sapiens cDNA clone IMAGE:1132411

DEFINITION similar to contains Alu repetitive element; contains element MER4

repetitive element ;, mRNA sequence.

ACCESSION AA631916

VERSION AA631916.1 GI:254527

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 380)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

Authors National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgaps-remail.nih.gov

Tissue Procurement: W. Marston Linehan, M.D., Rodrigo F. Chuagui, M.D.,

Michael Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: David B. Kitzman, Ph.D.

CDNA Library Arrayed by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.1nl.gov/bbrp/image/image.html
Insert Length: 470 Std Error: 0.00
Seq primer: -40ml3 fwd. ER from Amer sham.
Location/Qualifiers

FEATURES

source

1. .380
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1132411"
/clone_lib="NCI_CGAP_P12"
/sex="Male"
/dev_stage="45 years old"
/lab_host="DH10B"

/note="Vector: PAMP10; Site 1: NotI; Site 2: EcoRI; 1st strand cDNA was primed with oligo(dT)17 on 50 ng of DNase-treated, total cellular RNA obtained from 5,000-10,000 microdissected preneoplastic cells histologically-determined to be prostatic intraepithelial neoplasia 2 (PIN2) cells. Double-stranded cDNA was ligated to EcoRI adaptors, 5 cycles of PCR applied to the cDNA with an adaptor-specific primer, and the resulting PCR product subcloned into PAMP10 by the UDG-cloning method (Life Technologies). Average insert size is 600 bp. NOTE: Not directionally cloned. This library was constructed by David Krizman."

BASE COUNT 87 a 92 c 90 g 111 t

Query Match 80.4%; Score 222; DB 9; Length 380;
Best Local Similarity 100.0%; Pred. No. 4.9e-107;
Matches 222; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 TGCTTGCAAGGTCAGGCAAGCTGATTTGGTCCACCTTTGGCAGAGAACAGCGATG 114

DB 321 TGCTTGCAAGGTCAGGCAAGCTGATTTGGTCCACCTTTGGCAGAGAACAGCGATG 262

QY 115 TTGTCGCCCATTTCTCGATCAGACGCGCCCATTTACTACCTCCAGAGTGTCTTT 174

DB 261 TTGTCGCCCATTTCTCGATCAGACGCGCCCATTTACTACCTCCAGAGTGTCTTT 202

QY 175 CTCCTAATAAAGAAACATCTACTTGAACATCTACTGGGGAGACCGAGAGTATGTC 234

DB 201 CTCCTAATAAAGAAACATCTACTTGAACATCTACTGGGGAGACCGAGAGTATGTC 142

QY 235 TCAGCCTGTAATTCGGAATTCGGGAGCGCGAGCAGAGAG 276

DB 141 TCAGCCTGTAATTCGGAATTCGGGAGCGCGAGCAGAGAG 100

RESULT 6
A1557474 642 bp mRNA linear EST 09-AUG-1999

LOCUS A1557474 PT2.1_7_602.r tumor2 Homo sapiens cDNA 3', mRNA sequence.

DEFINITION A1557474

ACCESSION A1557474.1 GI:4489837

VERSION EST.

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 642)

Huang G.M., Ng,W.L., Farkas,J., He,L., Liang,H.A., Gordon,D., Yu,J.

and Hood,L.

Prostate cancer expression profiling by cDNA sequencing analysis

Genomics 59 (2), 178-186 (1999)

CONTACT: Guyang Matthew Huang

Leroy Hood

University of Washington

Department of Molecular Biotechnology, Box 357730, University of

Washington, Seattle, WA 98195

Tel: 5106280108

Fax: 5106280108

Email: huanggm@yahoo.com.

Location/Qualifiers

1. .642

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="tumor2"

/note="Organ: Prostate; Vector: pBluescript; Directional

cDNA library was constructed using Lambda Zp II kit

(Stratagene). mRNA was extracted from a frozen prostate

tumor tissue (Mayo Clinics)."

BASE COUNT 165 a 127 c 132 g 171 t 47 others

ORIGIN

Query Match 80.4%; Score 222; DB 9; Length 642;

Best Local Similarity 100.0%; Pred. No. 5.3e-107;

Matches 222; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 TGCTTGCAAGGTCAGGCAAGCTGATTTGGTCCACCTTTGGCAGAGAACAGCGATG 114

DB 38 TGCTTGCAAGGTCAGGCAAGCTGATTTGGTCCACCTTTGGCAGAGAACAGCGATG 97

QY 115 TTGTCGCCCATTTCTCGATCAGACGCGCCCATTTACTACCTCCAGAGTGTCTTT 174

DB 98 TTGTCGCCCATTTCTCGATCAGACGCGCCCATTTACTACCTCCAGAGTGTCTTT 157

QY 175 CTCCTAATAAAGAAACATCTACTTGAACATCTACTGGGGAGACCGAGAGTATGTC 234

DB 158 CTCCTAATAAAGAAACATCTACTTGAACATCTACTGGGGAGACCGAGAGTATGTC 217

QY 235 TCAGCCTGTAATTCGGAATTCGGGAGCGCGAGCAGAGAG 276

DB 218 TCAGCCTGTAATTCGGAATTCGGGAGCGCGAGCAGAGAG 259

RESULT 7
A1417931 294 bp mRNA linear EST 09-FEB-1999

LOCUS A1417931 tg55e07.x1 NCI_CGAP_P128 Homo sapiens cDNA clone IMAGE:2112708 3'

DEFINITION A1417931 similar to contains Alu repetitive element; contains element LTR3

ACCESSION A1417931

VERSION A1417931.1 GI:4261435

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 294)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgaabs-r@mail.nih.gov

Tissue procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/FLMI at:

www-bio.1nl.gov/bbrp/image/image.html

Seq primer: -40UP from Gluco.

Location/Qualifiers

1. .294

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:2112708"

/clone_lib="NCI_CGAP_P128"

/sex="male"

/dev_stage="adult"

/note="Organ: prostate; Vector: pT730-Pac (Pharmacia)

/lab_host="DH10B"

with a modified polylinker; Plasmid DNA from the

normalized library NCI-CGAP_P122 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clonoids 985608-986759, 1101192-1101959, and 1217928-1220615). Subtraction by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 71 a 63 c 66 g 94 t

Query Match 69.2%; Score 191; DB 9; Length 294;
Best Local Similarity 100.0%; Pred. No. 1.3e-90;
Matches 191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 TCCCCACCTTGCAGAGACAGCATGTGTGGCCCATTTCTCGATCAAGACCGG 145
DB 294 TCCCCACCTTGCAGAGACAGCATGTGTGGCCCATTTCTCGATCAAGACCGG 235
QY 146 CCCATCTTACTACTCCCAAGAGTGTCTTCTCTAATAAGAAAACATCTACTTTGAAC 205
DB 234 CCATCTTACTACTCCCAAGAGTGTCTTCTCTAATAAGAAAACATCTACTTTGAAC 175
QY 206 ATCTACTGGCGAGACAGAGTGTGTGCTGACCTGTAATTCTGAATTTGGGAGGCC 265
DB 174 ATCTACTGGCGAGACAGAGTGTGTGCTGACCTGTAATTCTGAATTTGGGAGGCC 115
QY 266 GAGCAGAGAG 276
DB 114 GAGCAGAGAG 104

RESULT 8
A1557019 760 bp mRNA linear EST 09-AUG-1999
LOCUS PT2.1_10_E08.r tumor2 Homo sapiens cDNA 3', mRNA sequence.
DEFINITION A1557019
ACCESSION A1557019
VERSION A1557019.1 GI:4489382
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS Huang, G.M., Ng, W.L., Farkas, J., He, L., Llang, H.A., Gordon, D., Yu, J. and Hood, L.
TITLE Prostate cancer expression profiling by cDNA sequencing analysis
JOURNAL Genomics 59 (2), 178-186 (1999)
MEDLINE 99339982
COMMENT Contact: Guyang Mathew Huang
Leroy Hood
University of Washington
Department of Molecular Biotechnology, Box 357730, University of Washington, Seattle, WA 98195
Tel: 5106280100
Fax: 5106280108
Email: huangm@yahoo.com.

FEATURES
source Location/Qualifiers

1..760
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="tumor2"
/note="Organ: Prostate; Vector: pBluescript; Directional cDNA library was constructed using lambda ZP II kit (Stratagene). mRNA was extracted from a frozen prostate tumor tissue (Mayo Clinics)."
BASE COUNT 176 a 146 c 164 g 195 t 79 others
ORIGIN

Query Match 68.5%; Score 189; DB 9; Length 760;
Best Local Similarity 100.0%; Pred. No. 1.8e-89;
Matches 189; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 88 CCCACCTTGCAGAGACAGCATGTGTGGCCCATTTCTCGATCAAGACCGGCC 147

DB 92 CCCACCTTGCAGAGACAGCATGTGTGGCCCATTTCTCGATCAAGACCGGCC 151

QY 148 CATCTTACTACTCCCAAGAGTGTCTTCTCTAATAAGAAAACATCTACTTTGAACAT 207
DB 152 CATCTTACTACTCCCAAGAGTGTCTTCTCTAATAAGAAAACATCTACTTTGAACAT 211

QY 208 CTACTGGCGAGACAGAGTGTGTGCTGACCTGTAATTCTGAATTTGGGAGGCCGA 267
DB 212 CTACTGGCGAGACAGAGTGTGTGCTGACCTGTAATTCTGAATTTGGGAGGCCGA 271

QY 268 GGCAGAGAG 276
DB 272 GGCAGAGAG 280

RESULT 9
AA876897 378 bp mRNA linear EST 25-MAR-1998
LOCUS ny47g12.s1 NCI-CGAP_P12 Homo sapiens cDNA clone IMAGE:1274950
DEFINITION AA876897
ACCESSION AA876897
VERSION AA876897.1 GI:2985974
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: W. Douglas Figg, Ph.D., Paul H. Durey, M.D., Rodrigo F. Chuquib, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Kitzman, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: www.bio.lnl.gov/bdrrp/image/image.html
Seq primer: -40m13 fwd. RT from Amersham
High quality sequence stop: 376.
Location/Qualifiers

1..378
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="IMAGE:1274950"
/clone_lib="NCI-CGAP_P12"
/sex="male"
/tissue_type="metastatic prostate bone lesion"
/lab_host="DH10B"
/note="Vector: pAMP10; mRNA made from metastatic prostate lesion of the bone, cDNA made by oligo-dT priming. Non-directionally cloned. Size-selected on agarose gel, average insert size 600 bp. Library made by D. Kitzman, NIH."

BASE COUNT 106 a 89 c 94 g 89 t
ORIGIN

Query Match 67.8%; Score 187; DB 9; Length 378;
Best Local Similarity 100.0%; Pred. No. 1.9e-88;
Matches 187; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 90 CACCTTGCAGAGACAGCATGTGTGGCCCATTTCTCGATCAAGACCGGCCCA 149
DB 98 CACCTTGCAGAGACAGCATGTGTGGCCCATTTCTCGATCAAGACCGGCCCA 157
QY 150 TCTTACTACTCCCAAGAGTGTGTGCTGACCTGTAATTCTGAATTTGAACATCT 209

Db 158 TCTTACTACCTCCAGAGATGCTTTCTCTCTATATAAGAAACATCTACTTTGAACATCT 217
QY 210 ACTGGCGCAGACACAGAGTGTAGTGCCTCAGCCCTGTAATTCCTGGAATTTGGAGCGCGAG 269
|||||
Db 218 ACTGGCGCAGACACAGAGTGTAGTGCCTCAGCCCTGTAATTCCTGGAATTTGGAGCGCGAG 277
QY 270 CAGGAAG 276
|||||
Db 278 CAGGAAG 284

RESULT 10
Bg612025 742 bp mRNA linear EST 18-APR-2001
LOCUS Bg612025
DEFINITION 602613927F1 NIH_MGC_60 Homo sapiens CDNA clone IMAGE:4739311 5',
mRNA sequence.
ACCESSION Bg612025
VERSION Bg612025.1 GI:13663396
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE

NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue Procurement: DCTD/DTP
CDNA Library Preparation: CLONETECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LMNL at:
http://image.llnl.gov

JOURNAL

COMMENT

High quality sequence stop: 369.
Location/Qualifiers
1. 742
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4739311"
/clone_id="NIH_MGC_60"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: prostate; Vector: pDNR-LIB (Clontech);
Site 1: SfiI (ggccgctggcc); Site 2: SfiI (ggccatattggcc
); Double-stranded CDNA was prepared from cell line RNA.
5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor
sequence: 5'-ATCTAGAGCGCCGAGCGCGCCAGT-dt(30)BN-3'
(where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.5 kb (range 0.9-4.0 kb). 14/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
library."

FEATURES

source

BASE COUNT 205 a 218 c 191 g 128 t
ORIGIN

Query Match 63.0%; Score 174; DB 10; Length 742;
Best Local Similarity 99.3%; Pred. No. 1.6e-81;
Matches 274; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTAAGGGCTGCAACAGAGCCCACTGGAGGCGCAACCTTTAGCGGATGCTTGC 60
|||||
Db 17 CTAAGGGCTGCAACAGAGCCCACTGGAGGCGCAACCTTTAGCGGATGCTTGC 76
|||||
QY 61 CAAAGTCAGCAGACCTGATTTCTGCTCCCACTTTGAGAGAGAGAGCATGTTGTGC 120
|||||
Db 77 CAAAGTCAGCAGACCTGATTTCTGCTCCCACTTTGAGAGAGAGCATGTTGTGC 136
|||||

QY 121 GCCCAATTTCTCAGATCAAGACCGGCCCATCTTACTACTCTCCAAAGATGCTTTCTCTCT 180
|||||
Db 137 GCCCAATTTCTCAGATCAAGACCGGCCCATCTTACTACTCTCCAAAGATGCTTTCTCTCT 196
|||||
QY 181 AATAAGAAACATCTACTTCTGAAACATCTACTGAGGCGAGACAGAGATGATGCTCAGCC 240
|||||
Db 197 AATAAGAAACATCTACTTCTGAAACATCTACTGAGGCGAGACAGAGATGATGCTCAGCC 256
|||||
QY 241 TGTAAATTCGTGAATTTCCGGAGCGCCGAGCAGAGAG 276
|||||
Db 257 TGTAAATTCGTGAATTTCCGGAGCGCCGAGCAGAGAG 292
|||||

RESULT 11
BF438241 297 bp mRNA linear EST 29-NOV-2000
LOCUS BF438241/C
DEFINITION 7901C02.X1 NCI CGAP Pr28 Homo sapiens cDNA clone IMAGE:3676394 3',
similar to contains Alu repetitive element.; mRNA sequence.
ACCESSION BF438241
VERSION BF438241.1 GI:11450758
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 297)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)

JOURNAL

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LMNL, send email to:
info@image.llnl.gov
Seq primer: -400P from Gibco.
Location/Qualifiers
1. 297
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3676394"
/clone_id="NCI_CGAP_Pr28"
/sex="male"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: prostate; Vector: p773D-Pac (Pharmacia)
with a modified polylinker; Plasmid DNA from the
normalized library NCI-CGAP Pr22 was prepared, and ss
circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (cloneIDs
985608-986759, 1101192-1101959, and 1217928-1220615).
Subtraction by Bento Soares and M. Fatima Bonaldo."

FEATURES

source

BASE COUNT 74 a 69 c 70 g 83 t
ORIGIN

Query Match 59.8%; Score 165; DB 10; Length 297;
Best Local Similarity 100.0%; Pred. No. 8.5e-77;
Matches 165; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 112 ATGTTGCGCCCATTTCTCAGATCAAGACCGGCCCATCTTACTACTCTCCAAAGATGCT 171
|||||
Db 253 ATGTTGCGCCCATTTCTCAGATCAAGACCGGCCCATCTTACTACTCTCCAAAGATGCT 194
|||||
QY 172 TTTCTCTCTAATAAGAAACATCTACTTGAACATCTACTGAGGCGAGACGAGATGAT 231
|||||
Db 193 TTTCTCTCTAATAAGAAACATCTACTTGAACATCTACTGAGGCGAGACGAGATGAT 134
|||||

QY 232 GGCCTCAGCCCTGTAATTCGTGAATTCGGAGCCGAGCAGAG 276
LOCUS |||||||
Db 133 GGCCTCAGCCCTGTAATTCGTGAATTCGGAGCCGAGCAGAG 89

RESULT 12
AA618586/c
LOCUS
DEFINITION np30h03.s1 NCI_CGAP_Pr22 Homo sapiens cDNA clone IMAGE:111787 3'
similar to contains Alu repetitive element; contains element LTR3
repetitive element ; mRNA sequence.

ACCESSION
VERSION AA618586
KEYWORDS
SOURCE EST. A618586.1 GI:2505791

ORGANISM human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www.bio.lnl.gov/bdrp/image/image.html
Insert Length: 424 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham.

FEATURES
source
location/Qualifiers
1..322
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone IMAGE:111787"
/clone_id="NCI_CGAP_Pr22"
/sex="male"
/tissue_type="normal prostate"
/lab_host="DH10B"
/note="Organ: prostate; Vector: pT73D-Pac (Pharmacia)
with a modified polylinker; 1st strand cDNA was prepared
from normal prostate bulk tissue, and was then primed with
a Not I - oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified pT73 vector. Library is normalized, and was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 80 a 77 c 76 g 89 t

ORIGIN
Query Match 47.5%; Score 131; DB 9; Length 322;
Best Local Similarity 100.0%; Pred. No. 9.7e-59;
Matches 131; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 TGCCTGCAAGCTCAGGAGCTGATCTGATCCCTTTCAGAGAGACAGCATG 114
LOCUS |||||||
Db 311 TGCCTGCAAGCTCAGGAGCTGATCTGATCCCTTTCAGAGAGACAGCATG 252

QY 115 TTGTGCGCCCATTTCTCAGATCAAGACCGGCCCATTTACTACCTCAAGAGTCTTTT 174
LOCUS |||||||
Db 251 TTGTGCGCCCATTTCTCAGATCAAGACCGGCCCATTTACTACCTCAAGAGTCTTTT 192

QY 175 CTCCTATAAA 185
LOCUS |||||||
Db 191 CTCCTATAAA 181

RESULT 13
BG498699
LOCUS
DEFINITION BG498699 551 bp mRNA linear EST 27-MAR-2001
602544315r1 NIH_MGC_60 Homo sapiens cDNA clone IMAGE:466923 5',
mRNA sequence.

ACCESSION
VERSION BG498699
KEYWORDS
SOURCE EST. BG498699.1 GI:13460216

ORGANISM human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS NIH-MGC http://mgc.ncbi.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCRD/DTP
CDNA Library Preparation: Clontech Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LICM1474 row: m column: 04
High quality sequence stop: 341.

FEATURES
source
location/Qualifiers
1..551
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone IMAGE:466923"
/clone_id="NIH_MGC_60"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: prostate; Vector: pDNR-LIB (Clontech);
site-1: SfiI (ggccgcctggcc); site-2: SfiI (ggccatattggc
); Double-stranded cDNA was prepared from cell line RNA.
5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CAAGGCGCATTAATGCGC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGCGCCGCGCGCGGACATG-dT(30)BN-3'
(where B = A, C, or G, or T). Average
insert size 1.5 kb (range 0.9-4.0 kb). 14/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
Library."

BASE COUNT 165 a 122 c 151 g 113 t

ORIGIN
Query Match 30.4%; Score 84; DB 10; Length 551;
Best Local Similarity 100.0%; Pred. No. 9.5e-34;
Matches 84; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 116 TTGTGCGCCCATTTCTCAGATCAAGACCGGCCCATTTACTACCTCAAGAGTCTTTT 175
LOCUS |||||||
Db 88 TTGTGCGCCCATTTCTCAGATCAAGACCGGCCCATTTACTACCTCAAGAGTCTTTT 147

QY 176 TCTCTAATAAGAAACATCTACTT 199
LOCUS |||||||
Db 148 TCTCTAATAAGAAACATCTACTT 171

RESULT 14
A1972706/c
LOCUS
DEFINITION A1972706 370 bp mRNA linear EST 25-AUG-1999
w142d04.x1 NCI CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2490343 3'
similar to contains Alu repetitive element; contains element MER4
repetitive element ; mRNA sequence.

ACCESSION
VERSION A1972706
KEYWORDS
SOURCE EST. A1972706.1 GI:5769532

human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 370)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Seq primer: -400P from Gibco.
Location/Qualifiers
1. 370
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2490343"
/clone_lib="NCI-CGAP_Pr28"
/sex="male"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: prostate; Vector: pT7/3D-Pac (pharmacia)
with a modified polylinker; Plasmid DNA from the
normalized library NCI-CGAP_Pr22 was prepared, and ss
circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (clonesids
985608-986759, 1101197-1101959, and 1217928-1220615).
Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 86 a 78 c 76 g 130 t
ORIGIN

Query Match 27.5%; Score 76; DB 9; Length 370;
Best Local Similarity 100.0%; Pred. No. 1.6e-29;
Matches 76; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 110 CGATGTTGCGCCCATTTCTCAGATCAAGACGGCCCATCTTACTACCTCCAGAGTG 169
|||||
Db 288 CGATGTTGCGCCCATTTCTCAGATCAAGACGGCCCATCTTACTACCTCCAGAGTG 229
|||||

QY 170 CTTTCTCTCTATAA 185
|||||
Db 228 CTTTCTCTCTATAA 213
|||||

RESULT 15
A0503931 545 bp DNA linear GSS 29-APR-1999
LOCUS RPCI-11-300D22.TV RPCI-11 Homo sapiens genomic clone RPCI-11-300D22
DEFINITION , DNA sequence.
ACCESSION A0503931
VERSION A0503931.1 GI:4708678
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 545)
AUTHORS Zhao, S., Adams, M.D., Niernan, W., Malek, J., de Jong, P. and Venter
J.C.
TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
Map Building
JOURNAL Unpublished (1997)
COMMENT Contact: Shaying Zhao, William Niernan, Mark Adams
Department of Eukaryotic Genomics

The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbeetfgr.org
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.bufileo.edu). Clones may be purchased from
BACPAC Resources (<http://bacpac.med.bufileo.edu/ordering>) or from
Research Genet cs (inforesgen.com). BAC end search page:
http://www.tigr.org/tcd/hungen/bac_end_search/bac_end_search.html.
Seq primer: T7
Class: BAC ends.
Location/Qualifiers
1. 545
/organism="Homo sapiens"
/db_xref="GDB:761490"
/db_xref="taxon:9606"
/clone="RPCI-11-300D22"
/clone_lib="RPCI-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBAC3.6; Site.1: EcoRI; Site.2: EcoRI;
RPCI11 Human Male BAC Library"
BASE COUNT 120 a 150 c 175 g 100 t
ORIGIN

Query Match 19.2%; Score 53; DB 12; Length 545;
Best Local Similarity 100.0%; Pred. No. 2.7e-17;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTAAGCGGTGCAAAACAGAGCGCCACTGGAGGCTGAACCTTAGGCCGATGC 53
|||||
Db 480 CTAAGCGGTGCAAAACAGAGCGCCACTGGAGGCTGAACCTTAGGCCGATGC 532
|||||

Search completed: October 8, 2002, 14:19:52
Job time : 1625 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 8, 2002, 12:11:47 : Search time 43 seconds
(without alignments)
1576.624 Million cell updates/sec

Title: US-09-065-672-5_COPY_1_276

Perfect score: 276
Sequence: 1 CTAAGCGGTGACAAACAGACG.....CGGAGAGCCGAGCAGGAG 276

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 383533 seqs, 122816752 residues

Word size : 0

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Issued_Patents_NA: *
1: /cgn2_6/ptodata/2/1na/5A.COMB.seq: *
2: /cgn2_6/ptodata/2/1na/5B.COMB.seq: *
3: /cgn2_6/ptodata/2/1na/6A.COMB.seq: *
4: /cgn2_6/ptodata/2/1na/6B.COMB.seq: *
5: /cgn2_6/ptodata/2/1na/PTUS.COMB.seq: *
6: /cgn2_6/ptodata/2/1na/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19	6.9	26664	4	US-09-564-805-28
2	19	6.9	49136	4	US-09-422-869-1
3	19	6.9	72604	4	US-09-268-992-7
4	17	6.2	688	1	US-08-599-252-94
5	17	6.2	688	5	PCT-US96-06352-94
6	17	6.2	688	5	PCT-US96-06583-94
7	17	6.2	3523	3	US-08-749-527-1
8	17	6.2	5359	3	US-09-070-060-1
9	17	6.2	5372	3	US-09-357-746-1
10	17	6.2	9421	4	US-09-224-834-2
11	17	6.2	9421	4	US-09-370-319C-2
12	17	6.2	11288	4	US-08-646-301A-1
13	17	6.2	11288	4	US-08-481-968A-4
14	17	6.2	11288	4	US-08-154-7128-4
15	17	6.2	22481	4	US-08-367-841A-43
16	17	6.2	22481	5	PCT-US95-07201-43
17	17	6.2	28720	4	US-09-341-587-7
18	17	6.2	59065	4	US-09-813-817-3
19	17	6.2	246240	2	US-08-724-394A-20
20	17	6.2	246240	2	US-08-724-394A-20
21	17	6.2	246240	2	US-08-724-394A-21
22	17	6.2	246240	2	US-08-724-394A-21
23	17	6.2	246240	2	US-08-724-394A-22
24	17	6.2	246240	2	US-08-724-394A-22
25	16	5.8	20	2	US-08-837-201C-25
26	16	5.8	20	4	US-09-364-416-25
27	16	5.8	301	4	US-09-439-313-299

28	16	5.8	577	4	US-09-227-357-92	Sequence 92, App1
29	16	5.8	619	4	US-09-385-982-358	Sequence 358, App
30	16	5.8	644	4	US-09-522-217-107	Sequence 107, App
31	16	5.8	674	4	US-09-328-111-76	Sequence 76, App1
32	16	5.8	799	3	US-09-095-485-1	Sequence 1, App1
33	16	5.8	821	4	US-08-352-902D-146	Sequence 146, App
34	16	5.8	926	4	US-08-938-669A-4	Sequence 4, App1
35	16	5.8	1260	1	US-08-599-252-83	Sequence 83, App1
36	16	5.8	1260	1	US-08-436-074-56	Sequence 56, App1
37	16	5.8	1260	5	PCT-US96-06352-83	Sequence 83, App1
38	16	5.8	1260	5	PCT-US96-06583-83	Sequence 83, App1
39	16	5.8	1297	6	5187077-23	Patent No. 5187077
40	16	5.8	1297	6	5427925-21	Patent No. 5427925
41	16	5.8	1418	5	PCT-US95-17111A-120	Sequence 120, App
42	16	5.8	1442	2	US-08-454-557C-120	Sequence 120, App
43	16	5.8	1442	2	US-08-340-426D-120	Sequence 120, App
44	16	5.8	1442	2	US-08-450-673C-120	Sequence 120, App
45	16	5.8	1566	4	US-08-482-073-7	Sequence 7, App1

ALIGNMENTS

RESULT 1
US-09-564-805-28
Sequence 28, Application US/09564805
Patent No. 6333403
GENERAL INFORMATION:
APPLICANT: Tavligian, Sean V.
APPLICANT: Teng, David H.F.
APPLICANT: Simard, Jacques
APPLICANT: Rommens, Johanna M.
APPLICANT: Myriad Genetics, Inc.
TITLE OF INVENTION: Chromosome 17p-Linked Prostate Cancer Susceptibility
FILE REFERENCE: 2318-258
CURRENT APPLICATION NUMBER: US/09/564,805
CURRENT FILING DATE: 2000-05-05
PRIOR APPLICATION NUMBER: US 60/107,468
PRIOR FILING DATE: 1998-11-06
PRIOR APPLICATION NUMBER: 09/434,382
PRIOR FILING DATE: 1999-11-05
NUMBER OF SEQ ID NOS: 240
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 28
LENGTH: 26664
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (910)..(13104)
OTHER INFORMATION: exon 1: 910-1154; exon 2: 1736-1786; exon 3:
OTHER INFORMATION: 1925-1995; exon 4: 3025-3089; exon 5: 4361-4418;
OTHER INFORMATION: exon 6: 5582-5650; exon 7: 7075-7194; exon 8:
OTHER INFORMATION: 8186-8244; exon 9: 12878-12936; exon 10:
NAME/KEY: misc_feature
LOCATION: (13756)..(22917)
OTHER INFORMATION: exon 11: 13756-13868; exon 12: 15283-15378; exon
OTHER INFORMATION: 13: 16278-16416; exon 14: 16498-16583; exon 15:
OTHER INFORMATION: 18583-18701; exon 16: 20349-20445; exon 17:
OTHER INFORMATION: 22172-22310; exon 18: 22879-22917
NAME/KEY: misc_feature
LOCATION: (23045)..(26452)
OTHER INFORMATION: exon 19: 23045-23154; exon 20: 23795-23895; exon
OTHER INFORMATION: 21: 23973-24093; exon 22: 24354-24432; exon 23:
OTHER INFORMATION: 25026-25170; exon 24: 25812-26036; polyadenylation
NAME/KEY: variation
LOCATION: (826)..(23879)
OTHER INFORMATION: s at positions 826 and 23180 is G or C; y at
OTHER INFORMATION: positions 1914, 5568, 7165, 16431, 1857 and 20486
OTHER INFORMATION: is C or T; n at position 13128 is t or tgat; r at
OTHER INFORMATION: positions 22211 and 23879 is A or G.

US-09-564-805-28

Query Match 6.9%; Score 19; DB 4; Length 26664;
Best Local Similarity 100.0%; Pred. No. 0.46;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 255 TCGGAGGCCGAGCAGCAGC 273

Db 18938 TCGGAGGCCGAGCAGCAGC 18956

RESULT 2

US-09-422-869-1/C

; Sequence 1, Application US/09422869
; Patent No. 6235481
; GENERAL INFORMATION:
; APPLICANT: POLONSKY, KENNETH S.
; APPLICANT: HORIKAWA, YUKIO
; APPLICANT: CDA, NAOHISA
; APPLICANT: COX, NANCY J.
; APPLICANT: SREENAN, SEAMUS
; APPLICANT: ZHOU, YUN-PING
; APPLICANT: OTANI, KENICHI
; APPLICANT: HANIS, CRAIG L.
; APPLICANT: BELL, GRAEME I.
; TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
; FILE REFERENCE: ARCD:307
; CURRENT APPLICATION NUMBER: US/09/422,869
; EARLIER FILING DATE: 1999-10-21
; EARLIER APPLICATION NUMBER: 60/134,175
; EARLIER FILING DATE: 1999-05-13
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 49136
; TYPE: DNA
; ORGANISM: Human
US-09-422-869-1

Query Match 6.9%; Score 19; DB 4; Length 49136;
Best Local Similarity 100.0%; Pred. No. 0.47;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 256 TCGGAGGCCGAGCAGCAGC 274

Db 25231 TCGGAGGCCGAGCAGCAGC 25213

RESULT 3

US-09-268-992-7

; Sequence 7, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
; APPLICANT: Chen, H.
; APPLICANT: Prelmer, N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; FILE REFERENCE: 7853-138
; CURRENT APPLICATION NUMBER: US/09/268,992
; EARLIER FILING DATE: 1999-03-16
; EARLIER APPLICATION NUMBER: 09/236,134
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 60/106,056
; EARLIER FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: 60/088,312
; EARLIER FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: 60/078,044
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 72604
; TYPE: DNA

; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: all n positions
; OTHER INFORMATION: n-a, c, g, or t
US-09-268-992-7

Query Match 6.9%; Score 19; DB 4; Length 72604;
Best Local Similarity 100.0%; Pred. No. 0.48;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 256 TCGGAGGCCGAGCAGCAGC 274

Db 28005 TCGGAGGCCGAGCAGCAGC 28023

RESULT 4

US-08-599-252-94

; Sequence 94, Application US/08599252
; Patent No. 5705343
; GENERAL INFORMATION:
; APPLICANT: DRAVNA, DENNIS T.
; APPLICANT: FEDER, JOHN N.
; APPLICANT: GRIERE, ANDREAS
; APPLICANT: KIMMEL, BRUCE E.
; APPLICANT: THOMAS, WINSTON J.
; APPLICANT: WOLFE, ROGER K.
; TITLE OF INVENTION: METHOD TO DIAGNOSE HEREDITARY
; TITLE OF INVENTION: HEMOCHROMATOSIS
; NUMBER OF SEQUENCES: 124
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Ave. N.W., Suite 5500
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1888
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/599,252
; FILING DATE: 09-FEB-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 9053-0001.21
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; TELEEX: 90-4030
; INFORMATION FOR SEQ ID NO: 94:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 688 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-599-252-94

Query Match 6.2%; Score 17; DB 1; Length 688;
Best Local Similarity 100.0%; Pred. No. 4.9;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 258 GGGAGGCCGAGCAGCAGC 274

Db 116 GGGAGGCCGAGCAGCAGC 132

RESULT 5

PCT-US96-06352-94

; Sequence 94, Application PC/TUS9606352
; GENERAL INFORMATION:
; APPLICANT: DRAVNA, DENNIS T.
; APPLICANT: FEDER, JOHN N.
; APPLICANT: GNIKE, ANDREAS
; APPLICANT: KIMMEL, BRUCE E.
; APPLICANT: THOMAS, WINSTON J.
; APPLICANT: WOLFE, ROGER K.
; TITLE OF INVENTION: METHOD TO DIAGNOSE HEREDITARY
; TITLE OF INVENTION: HEMOCHROMATOSIS
; NUMBER OF SEQUENCES: 124
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Ave. N.W., Suite 5500
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1888
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/06352
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/599,252
; FILING DATE: 09-FEB-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; TELEX: 90-4030
; INFORMATION FOR SEQ ID NO: 94:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 688 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; PCT-US96-06352-94

Query Match 6.2%; Score 17; DB 5; Length 688;
Best local similarity 100.0%; Pred. No. 4.9;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 258 GGGAGCGCGAGCAGGA 274
Db 116 GGGAGCGCGAGCAGGA 132

RESULT 6
PCT-US96-06583-94
; Sequence 94, Application PC/TUS9606583
; GENERAL INFORMATION:
; APPLICANT: DRAVNA, DENNIS T.
; APPLICANT: FEDER, JOHN N.
; APPLICANT: GNIKE, ANDREAS
; APPLICANT: KIMMEL, BRUCE E.
; APPLICANT: THOMAS, WINSTON J.
; APPLICANT: WOLFE, ROGER K.
; TITLE OF INVENTION: METHOD TO DIAGNOSE HEREDITARY
; TITLE OF INVENTION: HEMOCHROMATOSIS
; NUMBER OF SEQUENCES: 124
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Ave. N.W., Suite 5500
; CITY: Washington
; STATE: DC

; COUNTRY: USA
; ZIP: 20006-1888
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/06583
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/599,252
; FILING DATE: 09-FEB-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; TELEX: 90-4030
; INFORMATION FOR SEQ ID NO: 94:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 688 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; PCT-US96-06583-94

Query Match 6.2%; Score 17; DB 5; Length 688;
Best local similarity 100.0%; Pred. No. 4.9;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 258 GGGAGCGCGAGCAGGA 274
Db 116 GGGAGCGCGAGCAGGA 132

RESULT 7
US-08-749-527-1
; Sequence 1, Application US/08749527
; Patent No. 6054632
; GENERAL INFORMATION:
; APPLICANT: Reid, Marion E.
; TITLE OF INVENTION: METHOD OF MAKING MONOCLONAL
; TITLE OF INVENTION: ANTIBODIES USING POLYMORPHIC
; TITLE OF INVENTION: TRANSGENIC ANIMALS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hoffmann & Baron, LLP
; STREET: 350 Jericho Turnpike
; CITY: Jericho
; STATE: New York
; COUNTRY: USA
; ZIP: 11753
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/749, 527
; FILING DATE: 15-NO. 6054632-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: O'Dea, Sean W.
; REGISTRATION NUMBER: 37690
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 822-3550
; TELEFAX: (516) 822-3582
; INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 3523 nucleotides
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
US-08-749-527-1

Query Match 6.2%; Score 17; DB 3; Length 3523;
Best Local Similarity 100.0%; Pred. No. 5.2;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 80 TCTGTGTCCTCCACCTTT 96
|||||
DB 803 TCTGTGTCCTCCACCTTT 819

RESULT 8

US-09-070-060-1
; Sequence 1, Application US/09070060
; Patent No. 5976849
; GENERAL INFORMATION:
; APPLICANT: Hustad, Carolyn M.
; TITLE OF INVENTION: Human E3 Ubiquitin Protein
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ZENECA Pharmaceuticals, Inc.
; STREET: 1800 Concord Pike
; CITY: Wilmington
; STATE: DE
; COUNTRY: USA
; ZIP: 19850-5437
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/070,060
; FILING DATE: 30-APR-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/073,839
; FILING DATE: 05-FEB-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Higgins, Patrick H
; REGISTRATION NUMBER: 39,709
; REFERENCE/DOCKET NUMBER: PHM.70312
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302.886.4889
; TELEFAX: 302.886.8221
; TELETYPE:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5359 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-09-070-060-1

Query Match 6.2%; Score 17; DB 2; Length 5359;
Best Local Similarity 100.0%; Pred. No. 5.3;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 258 GGGAGCGCGAGCGAGCA 274
|||||
DB 3488 GGGAGCGCGAGCGAGCA 3504

RESULT 9

US-09-357-746-1
; Sequence 1, Application US/09357746
; Patent No. 6087122
; GENERAL INFORMATION:
; APPLICANT: ZENECA Limited
; TITLE OF INVENTION: HUMAN E3 UBIQUITIN PROTEIN LIGASE
; FILE REFERENCE: PHM.70312.N1
; CURRENT APPLICATION NUMBER: US/09/357,746
; EARLIER FILING DATE: 1999-07-21
; EARLIER APPLICATION NUMBER: US No. 6087122 60/073,839
; EARLIER FILING DATE: 1998-02-05
; EARLIER APPLICATION NUMBER: US No. 608712209/070,060
; EARLIER FILING DATE: 1998-04-30
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 5372
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-357-746-1

Query Match 6.2%; Score 17; DB 3; Length 5372;
Best Local Similarity 100.0%; Pred. No. 5.3;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 258 GGGAGCGCGAGCGAGCA 274
|||||
DB 3483 GGGAGCGCGAGCGAGCA 3499

RESULT 10

US-08-370-319C-2/c
; Sequence 2, Application US/08370319C
; Patent No. 5856091
; GENERAL INFORMATION:
; APPLICANT: Brichard, Vincent; Van Pel, Aline;
; APPLICANT: Traversari, Catia; W Ifel, Thomas; Coulie, Pierre;
; APPLICANT: Boon-Falleur, Thierry; De Plaen, Etienne
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID SEQUENCE CODING FOR A
; TITLE OF INVENTION: TUMOR REJECTION ANTIGEN PRECURSOR PROCESSED TO AT LEAST ONE
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felle & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/370,319C
; FILING DATE: 10-JANUARY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/272,351
; FILING DATE: 8-JULY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/032,978
; FILING DATE: 18-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5856091man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5377.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:

LENGTH: 9421 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

FEATURE:
OTHER INFORMATION: Following position there is an
OTHER INFORMATION: unsequenced portion of from 4.7 to 5.3
OTHER INFORMATION: kilobases
US-08-370-319C-2

Query Match 6.2%; Score 17; DB 2; Length 9421;
Best Local Similarity 100.0%; Pred. No. 5.4;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 258 GGGAGCCGAGCAGCA 274
Db 7499 GGGAGCCGAGCAGCA 7483

RESULT 11

US-09-224-834-2/c
Sequence 2, Application US/09224834
Patent No. 620111

GENERAL INFORMATION:
APPLICANT: Brichard, Vincent; Van Pel, Aline;
APPLICANT: Traversari, Catia; W lfel, Thomas; Pierre;
APPLICANT: Boon-Falleur, Thierry; De Plaen, Etienne
TITLE OF INVENTION: ISOLATED NUCLEIC ACID SEQUENCE CODING FOR A
TITLE OF INVENTION: TUMOR REJECTION ANTIGEN PRECURSOR PROCESSED TO AT LEAST ONE TU
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Felfe & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
COUNTRY: USA
ZIP: 10022

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/224,834
FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/370,319
FILING DATE:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/032,978
FILING DATE: 18-MAR-1993

ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 620111man D.
REGISTRATION NUMBER: 30,946

REFERENCE/DOCKET NUMBER: LUD 5377.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200

TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 9421 base pairs
TYPE: nucleic acid

STRANDEDNESS: double
TOPOLOGY: linear

FEATURE:

OTHER INFORMATION: Following position there is an
OTHER INFORMATION: unsequenced portion of from 4.7 to 5.3
OTHER INFORMATION: kilobases
US-09-224-834-2

Query Match 6.2%; Score 17; DB 4; Length 9421;

Best Local Similarity 100.0%; Pred. No. 5.4;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 258 GGGAGCCGAGCAGCA 274
Db 7499 GGGAGCCGAGCAGCA 7483

RESULT 12

US-08-646-301A-1
Sequence 1, Application US/08646301A
Patent No. 6194211

GENERAL INFORMATION:
APPLICANT: Richards, Cynthia Ann
APPLICANT: Huber, Brian E.
TITLE OF INVENTION: Transcriptional Regulatory Sequence of Carcinoembryonic
Patent No. 6194211
TITLE OF INVENTION: Antigen for Expression Targeting
FILE REFERENCE: PB1508USW
CURRENT APPLICATION NUMBER: US/08/646,301A
CURRENT FILING DATE: 1996-05-16
NUMBER OF SEQ ID NOS: 25
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 11288
TYPE: DNA
ORGANISM: Homo sapiens
US-08-646-301A-1

Query Match 6.2%; Score 17; DB 4; Length 11288;
Best Local Similarity 100.0%; Pred. No. 5.5;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 258 GGGAGCCGAGCAGCA 274
Db 6746 GGGAGCCGAGCAGCA 6762

RESULT 13

US-08-481-968A-4
Sequence 4, Application US/08481968A
Patent No. 6300490

GENERAL INFORMATION:
APPLICANT: Huber, Brian
APPLICANT: Richards, Cynthia
TITLE OF INVENTION: Molecular Constructs Comprising a Carcinoembryonic Antigen (

FILE REFERENCE: PB1087US4
CURRENT APPLICATION NUMBER: US/08/481,968A
CURRENT FILING DATE: 1998-06-07

NUMBER OF SEQ ID NOS: 36
SOFTWARE: PatentIn version 3.0
SEQ ID NO 4
LENGTH: 11288

TYPE: DNA
ORGANISM: Homo sapiens
US-08-481-968A-4

Query Match 6.2%; Score 17; DB 4; Length 11288;
Best Local Similarity 100.0%; Pred. No. 5.5;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 258 GGGAGCCGAGCAGCA 274
Db 6746 GGGAGCCGAGCAGCA 6762

RESULT 14

US-08-154-712B-4
Sequence 4, Application US/08154712B
Patent No. 6337209

GENERAL INFORMATION:
APPLICANT: Huber, Brian

```

; APPLICANT: Richards, Cynthia
; TITLE OF INVENTION: Molecular Constructs Containing a Carcinoembryonic Antigen Regu
; FILE OF INVENTION: Sequence
; FILE REFERENCE: P1087U53
; CURRENT APPLICATION NUMBER: US/08/154,712B
; CURRENT FILING DATE: 1993-11-19
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 11288
; TYPE: DNA
; ORGANISM: Homo sapiens
US-08-154-712B-4

Query Match
Best Local Similarity 100.0%; Pred. No. 5.5;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 258 GGGAGGCCGAGCAGCA 274
Db 6746 GGGAGGCCGAGCAGCA 6762
```

```

RESULT 15
US-08-367-841A-43/c
; Sequence 43, Application US/08367841A
; Patent No. 6319687
; GENERAL INFORMATION:
; APPLICANT: Chader, Gerald J.; Rodriguez,
; APPLICANT: Ignacio R.; Mazuruk, Krzysztof;
; APPLICANT: Tombran-Tink, Joyce
; TITLE OF INVENTION: PIGMENT EPITHELIUM
; TITLE OF INVENTION: DERIVED FACTOR: CHARACTERIZATION GENOMIC
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morgan & Flinnegan
; STREET: 345 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/367,841A
; FILING DATE: 30-DEC-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/257,963
; FILING DATE: 07-JUN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/952,796
; FILING DATE: 24-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: DOROTHY R. AUTH
; REGISTRATION NUMBER: 36434
; REFERENCE/DOCKET NUMBER: 20264126US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 22481 Base Pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Double
; TOPOLOGY: Unknown
; MOLECULE TYPE: Genomic DNA
; FEATURE:
; NAME/KEY: P1-147
```

```

; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION: full length genomic
; OTHER INFORMATION: sequence for PEDF plus flanking sequences.
US-08-367-841A-43

Query Match
Best Local Similarity 100.0%; Pred. No. 5.6;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 258 GGGAGGCCGAGCAGCA 274
Db 12071 GGGAGGCCGAGCAGCA 12055
```

Search completed: October 8, 2002, 12:27:03
Job time : 102 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 8, 2002, 12:23:32 : Search time 201 Seconds
(without alignments)
2357.534 Million cell updates/sec

Title: US-09-065-672-5_COPY_1_276

Perfect score: 276
Sequence: 1 CTAAGCGCTGCACAAACAGAC.....CGGAGCGCCGAGCAGAGAG 276

Scoring table:
Gapop 60.0 , Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

N_Geneseq_032802:*

1: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	276	100.0	343	21 AAC06853	Human secreted pro
2	176	63.8	230	AAF84110	Human cancer speci
3	132	47.8	660	AAF84111	Human cancer speci
4	21	7.6	21	AAF84117	Human CSG Prol1 s
5	20	7.2	20	AAF84116	Human CSG Prol1 s
6	20	7.2	375	AAI12572	Human breast cance
7	20	7.2	397	AAI18355	Human polynucleoti
8	20	7.2	528	AAI21445	Human breast cance
9	20	7.2	1151	AAI13351	Human secreted pro

c	10	20	7.2	1420	22	AAK75776	Human immune/haema
	11	20	7.2	19206	22	AAS35758	Human cardiovascular
	12	20	7.2	19274	22	AAS35754	Human cardiovascular
	13	20	7.2	24028	22	AAI05519	Human reproductive
	14	20	7.2	25574	22	AAI05619	Human reproductive
	15	20	7.2	25574	22	AAK79671	Human immune/haema
	16	20	7.2	25574	22	AAK83760	Human immune/haema
	17	20	7.2	25576	22	AAI05618	Human reproductive
	18	20	7.2	25576	22	AAK79669	Human immune/haema
	19	20	7.2	25576	22	AAK83758	Human immune/haema
	20	20	7.2	25576	22	AAK83305	Human immune/haema
	21	20	7.2	31348	22	AAS35759	Human cardiovascular
	22	20	7.2	31348	22	AAK79227	Human immune/haema
	23	20	7.2	50000	21	AAA96363	Polyomorph repeat
	24	20	6.9	51	22	AAI76313	Human silent SNP c
	25	19	6.9	89	21	AAC03085	Human secreted pro
	26	19	6.9	126	22	AAK83725	Human immune/haema
	27	19	6.9	126	22	AAK83727	Human immune/haema
	28	19	6.9	126	22	AAK83729	Human immune/haema
	29	19	6.9	132	22	AAK85587	Human immune/haema
	30	19	6.9	132	22	AAK85588	Human immune/haema
	31	19	6.9	138	22	AAK83723	Human immune/haema
	32	19	6.9	140	22	AAK79794	Human immune/haema
	33	19	6.9	145	21	AAC21835	Human secreted pro
	34	19	6.9	148	22	AAK62231	Human immune/haema
	35	19	6.9	150	22	AAI03446	Human reproductive
	36	19	6.9	158	16	AAI25057	Human gene signatu
	37	19	6.9	166	22	AAI03445	Human reproductive
	38	19	6.9	186	21	AAC17289	Human secreted pro
	39	19	6.9	216	21	AAC04787	Human secreted pro
	40	19	6.9	221	21	AAC03090	Human secreted pro
	41	19	6.9	221	21	AAC19819	Human secreted pro
	42	19	6.9	258	21	AAC06058	Human secreted pro
	43	19	6.9	265	22	ABA20116	Human nervous syst
	44	19	6.9	266	22	ABA16925	Human nervous syst
	45	19	6.9	276	21	AAC27411	Human secreted pro

ALIGNMENTS

RESULT 1
AAC06853
AAC06853 standard; cDNA: 343 BP.

06-OCT-2000 (first entry)

Human secreted protein 5' EST, SEQ ID NO: 10928.

Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
gene therapy; chromosome mapping; ss.

Homo sapiens.

EP1033401-A2.

06-SEP-2000.

21-FEB-2000; 2000EP-0200610.

26-FEB-1999; 99US-0122487.

(GEST) GENSET.

Dumas Milne Edwards J, Duclert A, Giordano J;
WPI; 2000-500381/45.

New nucleic acid that is a 5' expressed sequence tag (5' EST) for
obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX Claim 1; SEQ ID 10928; 71pp + CD-ROM; English.
PS
XX The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors.
XX
SQ Sequence 343 BP; 85 A; 90 C; 90 G; 77 T; 1 other;

Query Match 100.0%; Score 276; DB 21; Length 343;
Best Local Similarity 100.0%; Pred. No. 8e-137;
Matches 276; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTAAGGCGTGAACAGAGCGGCGACTGTGAACCTTTAGGCGATGCTGCTTG 60
DB 18 TTAAGGCGTGAACAGAGCGGCGACTGTGAACCTTTAGGCGATGCTGCTTG 77

QY 61 CAAGGTCAAGCAAGCTGTGTTGTCCTCCACTTTGACAGAGAGACGATGTTGTC 120
DB 78 CAAGGTCAAGCAAGCTGTGTTGTCCTCCACTTTGACAGAGAGACGATGTTGTC 137

QY 121 GCCCATTTTCAGATCAAGACCGGCCCATCTTACTCCTCAAGAGTCTTTCTCT 180
DB 138 GCCCATTTTCAGATCAAGACCGGCCCATCTTACTCCTCAAGAGTCTTTCTCT 197

QY 181 AATAGAAACATCTACTTGAACCTTACTGGGCGAGACGAGATGGCTAGGCC 240
DB 198 AATAGAAACATCTACTTGAACCTTACTGGGCGAGACGAGATGGCTAGGCC 257

QY 241 TGTAAATTCGTAATTCGGAGCGCGAGCAGAGAAAG 276
DB 258 TGTAAATTCGTAATTCGGAGCGCGAGCAGAGAAAG 293

RESULT 2
AAF84110
ID AAF84110 standard; DNA; 230 BP.
XX
AC AAF84110;
XX
DT 22-AUG-2001 (first entry)
XX
DE Human cancer specific gene (CSG) Pro121 EST sequence.
XX
KW Cancer specific gene; CSG; diagnostic marker; prostate; cancer; Pro119;
KW Pro121; Pro124; EST; expressed sequence tag; cytoslatic; gene therapy;
KW vaccine; ds.
XX
OS Homo sapiens.
XX
PN WO200139798-A1.
XX
PD 07-JUN-2001.
XX
PF 05-DEC-2000; 2000MO-US32927.
XX
PR 06-DEC-1999; 99US-0169083.
XX
PA (DIAD-) DIADEXUS INC.
XX
PI A11 S, Caferkey R, Recipon H, Sun Y;
XX

DR WPI; 2001-389934/41.
XX
XX Novel cancer specific gene and its protein useful for detecting,
PT diagnosing, monitoring, staging, prognosticating, imaging and treating
PT prostate cancer
XX
PS Claim 1; Page 43; 52pp; English.
XX
XX The invention relates to cancer specific genes (CSG) that have been
CC identified to be diagnostic markers for prostate cancer. The CSG genes
CC Pro119, Pro121 and Pro124 are useful as diagnostic markers for detecting,
CC diagnosing (metastases and disease), monitoring (cancer and changes in
CC cancer), staging, prognosticating, imaging and treating prostate cancer.
CC The CSG protein is useful for inducing an immune response against target
CC cell expressing a CSG. The present sequence represents an EST (expressed
CC sequence tag) sequence for the CSG Pro121.
XX
SQ Sequence 230 BP; 56 A; 58 C; 57 G; 58 T; 1 other;

Query Match 63.8%; Score 176; DB 22; Length 230;
Best Local Similarity 100.0%; Pred. No. 1.2e-83;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 TGGTTCAGAGTCAAGCAAGCTGATTCGTGTCCTCCACCTTTGACAGAGAACAGGATG 114
DB 28 TGGTTCAGAGTCAAGCAAGCTGATTCGTGTCCTCCACCTTTGACAGAGAACAGGATG 87

QY 115 TTGTGGCCCATTTTCAGATCAAGACCGGCCCATCTTACTACCTCAAGAGTCTTTT 174
DB 88 TTGTGGCCCATTTTCAGATCAAGACCGGCCCATCTTACTACCTCAAGAGTCTTTT 147

QY 175 CTCTTAATTAAGAAACATCTACTTTGAACATCTCTGCGGAGACCGAGAGTGA 230
DB 148 CTCTTAATTAAGAAACATCTACTTTGAACATCTCTGCGGAGACCGAGAGTGA 203

RESULT 3
AAF84111
ID AAF84111 standard; DNA; 660 BP.
XX
AC AAF84111;
XX
DT 22-AUG-2001 (first entry)
XX
DE Human cancer specific gene (CSG) Pro121 sequence.
XX
KW Cancer specific gene; CSG; diagnostic marker; prostate; cancer; Pro119;
KW Pro121; Pro124; EST; expressed sequence tag; cytoslatic; gene therapy;
KW vaccine; ds.
XX
OS Homo sapiens.
XX
PN WO200139798-A1.
XX
PD 07-JUN-2001.
XX
PF 05-DEC-2000; 2000MO-US32927.
XX
PR 06-DEC-1999; 99US-0169083.
XX
PA (DIAD-) DIADEXUS INC.
XX
PI A11 S, Caferkey R, Recipon H, Sun Y;
XX
DR WPI; 2001-389934/41.
XX
PT Novel cancer specific gene and its protein useful for detecting,
PT diagnosing, monitoring, staging, prognosticating, imaging and treating
PT prostate cancer
XX
PS Claim 1; Page 43; 52pp; English.
XX
XX The invention relates to cancer specific genes (CSG) that have been

CC identified to be diagnostic markers for prostate cancer. The CSG genes
CC Prol19, Prol21 and prol24 are useful as diagnostic markers for detecting
CC diagnosing (metastases and disease), monitoring (cancer and changes in
CC cancer), staging, prognosticating, imaging and treating prostate cancer.
CC The CSG protein is useful for inducing an immune response against target
CC cell expressing a CSG. The present sequence represents the CSG Prol21
CC sequence.
XX
XX
XX Sequence 660 BP; 172 A; 148 C; 190 G; 150 T; 0 other;

Query Match	47.8%;	Score 132;	DB 22;	Length 660;
Best Local Similarity	100.0%;	Pred. No. 3e-60;		
Matches 132;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy	55	TGTTGCAAGGTCAGGCAAGCTGATTCGTGGTCCCACTTTGGAGAGAGAACAGCGATG	114
	79	TGCTTTCGAAAGTCAGGCAAGCTGATTCGTGGTCCCACTTTGGAGAGAGAACAGCGATG	138
Qy	115	TTTGCGCGCCATTTCTGATCAAGGACCGGCCCATCTTACTACCTCAGAGAGTCCTTT	174
	139	TTTGCGCGCCATTTCTGATCAAGGACCGGCCCATCTTACTACCTCAGAGAGTCCTTT	198
Db			
Qy	175	CTCTCTAATAAG	186
	199	CTCTCTAATAAG	210
Db			

RESULT 4
AAF84117/c
ID AAF84117 standard; DNA; 21 BP.

AC	AAF84117;
XX	.
DT	22-AUG-2001 (first entry)
DT	

Human CSG Pro121 specific reverse primer.

KM Cancer specific gene; CSG; diagnostic marker; prostate; cancer; Prol19,
KM Prol21; Prol24; EST; expressed sequence tag; cytostatic; gene therapy;
KM vaccine; PCR primer; ss.

OS Homo sapiens.

PN WO200139798-A1.

PD 07-JUN-2001.

PF 05-DEC-2000; 2000WO-US32927.

PR 06-DEC-1999; 99US-0169083.

PA (DIAD-) DIADEXUS INC.

PI All S, Caferkey R, Recipon H, Sun Y;

DR WPI; 2001-389934/41.

PT	Novel cancer specific gene and its protein useful for detecting,
PT	diagnosing, monitoring, staging, prognosticating, imaging and treating
PT	prostate cancer -

PS Examples; Page 25; 52pp; English.

CC The invention relates to cancer specific genes (CSG) that have been
CC identified to be diagnostic markers for prostate cancer. The CSG genes
CC Prol19, Prol21 and Prol24 are useful as diagnostic markers for detecting
CC diagnosing (metastases and disease), monitoring (cancer and changes in
CC cancer), staging, prognosticating, imaging and treating prostate cancer.
CC The CSG protein is useful for inducing an immune response against target
CC cell expressing a CSG. Sequences AAF84116-117 represent primers used in
CC real-time quantitative PCR analysis of CSG Prol21.

Sequence 21 BP; 6 A; 5 C; 5 G; 5 T; 0 other;

Query Match	Score 21;	DB 22;	Length 21;
Best Local Similarity	100.0%;	Pred. No. 0.31;	
Matches	21; Conservative	0; Mismatches	0; Indels
			Gaps 0;
QY	137	CTTTGAACATCTACTGCGCG	217
Db	21	CTTTGAACATCTACTGCGCG	1

RESULT 5
AAF84116
ID AAF84116 standard; DNA; 20 BP.

AC	AAF84116;
XX	
DT	22-AUG-2001 (first entry)

Human CSG Prol21 specific forward primer.

KM Cancer specific gene; CSG; diagnostic marker; prostate; cancer; Prol19;
 KM Prol21; Prol24; EST; expressed sequence tag; cytostatic; gene therapy;
 KM vaccine; PCR primer; ss.

OS Homo sapiens.

PN W0200139798-A1.

PD 07-JUN-2001.

PF 05-DEC-2000; 2000WO-US32927.

PR 06-DEC-1999; 99US-0169083.

PA (DIAD-) DIADEXUS INC.

PI Ali S, Caferkey R, Recipon H, Sun Y;

DR WPI; 2001-389934/41.

PT Novel cancer specific gene and its protein useful for detecting,
PT diagnosing, monitoring, staging, prognosticating, imaging and treating
PT prostate cancer -

PS Examples; Page 25; 52pp; English.

CC The Invitron relates to cancer specific genes (CSG) that have been
CC identified to be diagnostic markers for prostate cancer. The CSG genes
CC Pro1p, Pro12l and Pro124 are useful as diagnostic markers for detecting
CC diagnosing (metastases and disease), monitoring (cancer and changes in
CC cancer), staging, prognosticating, imaging and treating prostate cancer.
CC The CSG protein is useful for inducing an immune response against target
CC cell-expressing a CSG. Sequences ARF84115-117 represent primers used in
CC real-time quantitative PCR analysis of CSG Pro12l.

SQ Sequence 20 BP; 5 A; 7 C; 3 G; 5 T; 0 other;

Query Match	7.2%	Score 20:	DB 22;	length 20;
Best Local Similarity	100.0%	Pred. No. 1;		
Matches 20; Conservative	0;	Mismatches	0;	Gaps 0;

QY 120 CGCCCATTTCTCAGATCAAG 139
|||||
Db 1 CGCCCATTTCTCAGATCAAG 20

RESULT 6

ID ALL12572 standard; cDNA; 375 BP

AC AAL12572;

DT 07-DEC-2001 (first entry)

XX DE Human breast cancer expressed polynucleotide 5029.
XX XX Human; breast cancer; cell marker; cytostatic; ss.
XX KW Homo sapiens.
XX OS WO200151628-A2.
XX PN 19-JUL-2001.
XX PD 10-JAN-2001; 2001WO-US00798.
XX PE 14-JAN-2000; 2000US-0176077.
XX PR 14-MAR-2000; 2000US-0189167.
XX PR 24-MAR-2000; 2000US-0192099.
XX PR 29-MAR-2000; 2000US-0193480.
XX PR 15-MAY-2000; 2000US-0205230.
XX PR 09-JUN-2000; 2000US-0211315.
XX PR 25-JUL-2000; 2000US-0220534.
XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX DR WPI; 2001-451856/48.
XX PT New peptide useful as a marker for the diagnosis of breast cancer -
XX PS Claim 1; Page 901; 3695pp; English.
XX CC The invention relates to human breast cancer expressed polynucleotides
XX CC (AAU07544-AAU26789) and methods of assessing whether a patient is
XX CC afflicted with breast cancer by examining the correlation between the
XX CC expression of certain markers and the cancerous state of breast cells.
XX CC The polynucleotides and encoded polypeptides are potential markers for
XX CC detecting, diagnosing, monitoring, characterising, treating and encoded
XX CC potentially preventing breast cancer. The polynucleotides and encoded
XX CC polypeptides are also useful for isolating compounds with cytostatic
XX CC activity.
SQ Sequence 375 BP; 113 A; 69 C; 107 G; 85 T; 1 other;
Query Match 7.2%; Score 20; DB 22; Length 375;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 255 TTCGGAGGCGCAGCAGCA 274
DB 195 TTCGGAGGCGCAGCAGCA 214
RESULT 7
AAI83855
ID AAI83855 standard; cDNA; 397 BP.
XX AC AAI83855;
XX AC
XX DT 06-NOV-2001 (first entry)
XX DE Human polynucleotide SEQ ID NO 3915.
XX DE
XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX KW vaccine; peptide therapy; stem cell growth factor; hematopoiesis;
XX KW tissue growth factor; immunomodulatory; cancer; leukaemia;
XX KW nervous system disorders; arthritis; inflammation; ss.
XX OS Homo sapiens.
XX PN WO200164835-A2.
XX PD 07-SEP-2001.
XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

PF 26-FEB-2001; 2001WO-US04927.
XX XX 28-FEB-2000; 2000US-0515126.
XX PR 18-MAY-2000; 2000US-0577409.
XX XX
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YF, Liu C, Drmanac RT;
XX DR WPI; 2001-514838/56.
XX DR P-PSDB; AAO03924.
XX XX
XX PT Isolated nucleic acids and polypeptides, useful for preventing
XX PT diagnosing and treating e.g. leukaemia, inflammation and immune
XX PT disorders -
XX PS Claim 1; SEQ ID NO 3915; 1399pp + Sequence Listing; English.
XX CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
XX CC the encoded proteins (AAO00010-AAO13910) that exhibit actively elating to
XX CC cytokine, cell proliferation or cell differentiation or which may induce
XX CC production of other cytokines in other cell populations. The
XX CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX CC peptide therapy. The polypeptides have various cytokine-like activities,
XX CC e.g. stem cell growth factor activity, hematopoiesis regulating
XX CC activity, tissue growth factor activity, immunomodulatory activity and
XX CC activin/inhibin activity and may be useful in the diagnosis and/or
XX CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX CC inflammation.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from Wiro
XX CC at ftp.wipo.int/pub/published_pct_sequences.
SQ Sequence 397 BP; 113 A; 76 C; 121 G; 87 T; 0 other;
Query Match 7.2%; Score 20; DB 22; Length 397;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 256 TCGGGAGGCGCAGCAGCA 275
DB 6 TCGGGAGGCGCAGCAGCA 25
RESULT 8
AAI21445
ID AAI21445 standard; cDNA; 528 BP.
XX AC AAI21445;
XX AC
XX DT 07-DEC-2001 (first entry)
XX DE Human breast cancer expressed polynucleotide 13902.
XX DE
XX KW Human; breast cancer; cell marker; cytostatic; ss.
XX KW
XX OS Homo sapiens.
XX PN WO200151628-A2.
XX PD 19-JUL-2001.
XX PF 10-JAN-2001; 2001WO-US00798.
XX PF
XX PR 14-JAN-2000; 2000US-0176077.
XX PR 14-MAR-2000; 2000US-0189167.
XX PR 24-MAR-2000; 2000US-0192099.
XX PR 29-MAR-2000; 2000US-0193480.
XX PR 15-MAY-2000; 2000US-0205230.
XX PR 09-JUN-2000; 2000US-0211315.
XX PR 25-JUL-2000; 2000US-0220534.
XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX Lillie J, Xu Y, Wang Y, Steimann K;
 PI WPI; 2001-451856/48.
 DR
 XX
 PT New peptide useful as a marker for the diagnosis of breast cancer -
 PS Claim 1; Page 2476; 3695pp; English.
 XX
 CC The invention relates to human breast cancer expressed polynucleotides
 CC (AA07544-AA07589) and methods of assessing whether a patient is
 CC afflicted with breast cancer by examining the correlation between the
 CC expression of certain markers and the cancerous state of breast cells.
 CC The polynucleotides and encoded polypeptides are potential markers for
 CC detecting, diagnosing, monitoring, characterising treating and
 CC potentially preventing breast cancer. The polynucleotides and encoded
 CC polypeptides are also useful for isolating compounds with cytostatic
 CC activity.
 CC
 SQ Sequence 528 BP; 168 A; 100 C; 145 G; 114 T; 1 other;
 Query Match 7.2%; Score 20; DB 22; Length 528;
 Best Local Similarity 100.0%; Pred. No. 1.1;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 255 TTCGGAGCGCGAGCAGCA 274
 ||||||||||||||||
 Db 222 TTCGGAGCGCGAGCAGCA 241
 RESULT 9
 AAD13351/c
 ID AAD13351 standard; cDNA; 1151 BP.
 AC AAD13351;
 XX
 DT 16-OCT-2001 (first entry)
 XX
 DE Human secreted protein-encoding gene 7 cDNA clone HTT131, SEQ ID NO:17.
 XX
 KW Human secreted protein; proliferative disorder; cancer; tumour;
 KW foetal abnormality; developmental abnormality; haematopoietic disorder;
 KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
 KW inflammation; allergy; neurological disorder; Alzheimer's disease;
 KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
 KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
 KW cardiovascular disorder; angiogenic disorder; kidney disorder;
 KW gastrointestinal disorder; pregnancy-related disorder;
 KW endocrine disorder; infection; wound healing; vulnerability;
 KW cell culture; chemotaxis; food additive; gene therapy;
 KW binding partner identification; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 29..313
 FT /*tag= a
 FT /product= "Human secreted protein precursor"
 FT sig_peptide 29..79
 FT /*tag= b
 FT mat_peptide 80..310
 FT /*tag= c
 FT /product= "Mature human secreted protein"
 XX
 PN WO200154708-A1.
 XX
 PD 02-AUG-2001.
 XX
 PE 17-JAN-2001; 2001WO-US01434.
 XX
 XX 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 18-AUG-2000; 2000US-0226279.

PR 05-DEC-2000; 2000US-0251988.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX Rosen CA, Komatsoulis GA, Baker KP, Birse CE, Soppet DR, Olsen HS;
 PI Moore PA, Wei P, Ebner R, Duan DR, Shi Y, Choi GH, Fiscella M;
 PI Ni J, Ruben SM, Barash SC;
 XX WPI; 2001-486743/53.
 DR P-PSDB; AA07057.
 DR
 XX
 PT New isolated nucleic acids and polypeptides, useful for diagnosing,
 PT treating and/or preventing human diseases and disorders -
 XX
 PS Claim 1; Page 451; 558pp; English.
 XX
 CC AAD13345-AA13401 represent cDNAs corresponding to 22 human secreted
 CC protein genes, and AA07051-AA07105 represent the proteins they encode.
 CC AA07106-AA07129 represent human secreted protein fragments or variants.
 CC
 CC The genes and their secreted proteins are useful for preventing,
 CC treating or ameliorating medical conditions, e.g., by protein or gene
 CC therapy. Pathological conditions can be diagnosed by determining the
 CC amount of the new protein in a sample or by determining the presence of
 CC mutations in the new genes. Specific uses are described for each of the
 CC 22 genes, based on the tissues in which they are most highly expressed,
 CC and include developing products for the diagnosis or treatment of
 CC proliferative disorders, cancer, tumours, foetal and developmental
 CC abnormalities, haematopoietic disorders, diseases of the immune system,
 CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
 CC allergies, neurological disorders (e.g., Alzheimer's disease,
 CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,
 CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
 CC cardiovascular disorders, angiogenic disorders, kidney disorders,
 CC gastrointestinal disorders, pregnancy-related disorders, endocrine
 CC disorders, and infections. The proteins can also be used to aid wound
 CC healing and epithelial cell proliferation, to prevent skin aging due to
 CC sunburn, to maintain organs before transplantation, for supporting cell
 CC culture of primary tissues, to regenerate tissues, to identify their
 CC cognate ligands or binding partners, and in chemotaxis, and can be used
 CC as a food additive or preservative to modify storage properties.
 CC Antibodies specific for a protein of the invention can be used in
 CC alleviating symptoms associated with the disorders mentioned above, and
 CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
 CC immunosorbent assay (ELISA). The present sequence represents a human
 CC secreted protein-encoding cDNA of the invention.
 CC
 SQ Sequence 1151 BP; 288 A; 306 C; 299 G; 258 T; 0 other;
 Query Match 7.2%; Score 20; DB 22; Length 1151;
 Best Local Similarity 100.0%; Pred. No. 1.1;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 255 TTCGGAGCGCGAGCAGCA 274
 ||||||||||||||||
 Db 442 TTCGGAGCGCGAGCAGCA 423
 RESULT 10
 AAK75776/c
 ID AAK75776 standard; DNA; 1420 BP.
 AC AAK75776;
 XX
 DT 07-NOV-2001 (first entry)
 XX
 DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:30588.
 XX
 KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
 KW cytostatic; gene therapy; vaccine; metastasis; ds.
 XX
 OS Homo sapiens.
 XX

PN WO200157182-A2.
XX 09-AUG-2001.
XX 17-JAN-2001; 2001WO-US01354.
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.

PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 06-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0255678.
(HUMA-) HUMAN GENOME SCI INC.
PA Rosen CA, Barash SC, Ruben SM;
XX
PI

XX WP1: 2001-483426/52.

XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,

XX useful for preventing, diagnosing and/or treating cancers and

XX metastasis -

PS Disclosure; SEQ ID NO 30588; 3071bp + Sequence Listing; English.

XX
XX
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)

CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic

CC activity, and can be used in gene therapy and vaccine production. (I)

CC proteins and polynucleotides may be used in the prevention, diagnosis and

CC treatment of diseases associated with inappropriate (I) expression. For

CC example, they may be used to treat disorders associated with decreased

CC expression by rectifying mutations or deletions in a patient's genome

CC that affect the activity of (I) by expressing inactive proteins or to

CC supplement the patient's own production of (I). Additionally, (I)

CC polynucleotides may be used to produce the secreted (I), by inserting

CC the nucleic acids into a host cell and culturing the cell to express the

CC protein. (I) proteins and polynucleotides may be used to prevent,

CC diagnose and treat immune/hematopoietic-related diseases, especially

CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703

CC to AAK87654 represent human immune/hematopoietic antigen genomic

CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169

CC represent sequences used in the exemplification of the present invention.

XX
XX
XX Sequence 1420 BP; 353 A; 320 C; 366 G; 381 T; 0 other.

Query Match 7.28; Score 20; DB 22; Length 1420;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 256 TCGGAGGCCGCGAGCAGGAA 275
|||||
Db 911 TCGGAGGCCGCGAGCAGGAA 892

RESULT 11
AAS35758
ID AAS35758 standard; DNA; 19206 BP.

XX
XX AAS35758;
XX
XX 17-DEC-2001 (first entry)

DE Human cardiovascular system antigen genomic DNA SEQ ID NO 1258.

XX
XX Cardiovascular system antigen; human; mouse; rabbit; goat; horse; cat;
XX chicken; sheep; immunosuppressive; antiarthritic; vasotropic; dog;
XX antirheumatic; antiproliferative; cyrostatic; cardiant; neuroprotective;
XX cerebroprotective; nootropic; antibacterial; virucide; fungicide; cancer;
XX ophtalmological; valnetary; gene therapy; autoimmune disease; neoplasm;
XX hyperproliferative disorder; breast; liver; cardiovascular disorder; ds;
XX cerebrovascular disorder; nervous system disorder; bacterial infection;
XX fungal infection; viral infection; ocular disorder; endocrine disorder;
XX gastrointestinal disorder; renal disorder; respiratory disorder;
XX wound healing; skin aging; organ transplantation; tissue regeneration;
XX anti-infertility.

XX
XX
XX Homo sapiens.
XX
XX
XX WO200155321-A2.
XX
XX
XX 02-AUG-2001.
XX
XX
XX 17-JAN-2001; 2001WO-US01340.
XX
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.

PR	13-OCT-2000;	2000US-0239937.
PR	20-OCT-2000;	2000US-0240960.
PR	20-OCT-2000;	2000US-0241221.
PR	20-OCT-2000;	2000US-0241785.
PR	20-OCT-2000;	2000US-0241786.
PR	20-OCT-2000;	2000US-0241787.
PR	20-OCT-2000;	2000US-0241808.
PR	20-OCT-2000;	2000US-0241809.
PR	20-OCT-2000;	2000US-0241826.
PR	01-NOV-2000;	2000US-0244617.
PR	08-NOV-2000;	2000US-0246474.
PR	08-NOV-2000;	2000US-0246475.
PR	08-NOV-2000;	2000US-0246476.
PR	08-NOV-2000;	2000US-0246477.
PR	08-NOV-2000;	2000US-0246478.
PR	08-NOV-2000;	2000US-0246523.
PR	08-NOV-2000;	2000US-0246524.
PR	08-NOV-2000;	2000US-0246525.
PR	08-NOV-2000;	2000US-0246526.
PR	08-NOV-2000;	2000US-0246527.
PR	08-NOV-2000;	2000US-0246528.
PR	08-NOV-2000;	2000US-0246532.
PR	08-NOV-2000;	2000US-0246609.
PR	08-NOV-2000;	2000US-0246610.
PR	08-NOV-2000;	2000US-0246611.
PR	08-NOV-2000;	2000US-0246613.
PR	17-NOV-2000;	2000US-0249207.
PR	17-NOV-2000;	2000US-0249208.
PR	17-NOV-2000;	2000US-0249209.
PR	17-NOV-2000;	2000US-0249210.
PR	17-NOV-2000;	2000US-0249211.
PR	17-NOV-2000;	2000US-0249212.
PR	17-NOV-2000;	2000US-0249213.
PR	17-NOV-2000;	2000US-0249214.
PR	17-NOV-2000;	2000US-0249215.
PR	17-NOV-2000;	2000US-0249216.
PR	17-NOV-2000;	2000US-0249217.
PR	17-NOV-2000;	2000US-0249218.
PR	17-NOV-2000;	2000US-0249244.
PR	17-NOV-2000;	2000US-0249245.
PR	17-NOV-2000;	2000US-0249264.
PR	17-NOV-2000;	2000US-0249265.
PR	17-NOV-2000;	2000US-0249297.
PR	17-NOV-2000;	2000US-0249299.
PR	17-NOV-2000;	2000US-0249300.
PR	01-DEC-2000;	2000US-0250160.
PR	01-DEC-2000;	2000US-0250391.
PR	05-DEC-2000;	2000US-0251030.
PR	05-DEC-2000;	2000US-0251988.
PR	05-DEC-2000;	2000US-0251989.
PR	06-DEC-2000;	2000US-0251989.
PR	06-DEC-2000;	2000US-0251989.
PR	08-DEC-2000;	2000US-0251868.
PR	08-DEC-2000;	2000US-0251869.
PR	08-DEC-2000;	2000US-0251869.
PR	11-DEC-2000;	2000US-0254097.
PR	05-JAN-2001;	2001US-0259678.
PA	(HUMA-) HUMAN GENOME SCI INC.	
XX		
XX		
PI	Rosen CA, Barash SC, Ruben SM;	
XX		
XX	WPI; 2001-451930/48.	
DR		
XX		
PT	New cardiovascular system related polynucleotides and polypeptides,	
PT	useful for diagnosing, treating and/or preventing disorders of the	
PT	cardiovascular system -	
XX		
PS	Claim 1; SEQ ID No 1258; 674pp; English.	
XX		
XX		
CC	Sequences AAS3741-AMS3642 represent genomic DNA molecules, which encode	
CC	the cardiovascular system antigen polypeptides of the invention.	

PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.

PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249266.
PR 17-NOV-2000; 2000US-0249267.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
WPI; 2001-451930/48.

New cardiovascular system related polynucleotides and polypeptides,
useful for diagnosing, treating and/or preventing disorders of the
cardiovascular system -
Claim 1; SEQ ID NO 1254; 674pp; English.

XX Sequences AAS35741-AAS36942 represent genomic DNA molecules, which encode
CC the cardiovascular system antigen polypeptides of the invention.
CC Cardiovascular system antigens and their associated polynucleotides are
CC useful in the diagnosis, treatment and prevention of various types of
CC disorders in e.g. humans, mice, rabbits, goats, horses, cats, dogs,
CC chickens or sheep. A pathological condition can be determined by
CC detecting the presence or absence of a mutation in a cardiovascular
CC system antigen polynucleotide. The treatable disorders include autoimmune
CC diseases such as rheumatoid arthritis, hyperproliferative disorders such

CC as neoplasms of the breast or liver, cardiovascular disorders such as
CC cardiac arrest, cerebrovascular disorders such as cerebral ischemia,
CC nervous system disorders such as Alzheimer's disease, infections caused
CC by bacteria, viruses and fungi, ocular disorders such as corneal
CC infection, endocrine disorders such as premature labour and infertility,
CC gastrointestinal disorders such as Crohn's disease, renal disorders such
CC as glomerulonephritis and respiratory disorders such as asthma and
CC pleurisy. The polypeptides can also be used to aid wound healing, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, to regenerate tissues and in chemotaxis.
CC Note: The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format directly from WIPO
CC at [ftp.wipo.int/pub/published_pcl_sequences](http://wipo.int/pub/published_pcl_sequences).
XX

Query Match 7.2%; Score 20; DB 22; Length 19274;
Best Local Similarity 100.0%; Pred. No. 1,2; Mismatches 0; Indels 0; Gaps 0;
Matches 20; Conservative 0;

OY 255 TTCGGAGGCCGAGCAGCA 274
DB 3956 TTCGGAGGCCGAGCAGCA 3975
|||||

RESULT 13

ID AAL05519 standard; DNA; 24028 BP.

XX AAL05519;

DT 21-NOV-2001 (first entry)

XX Human reproductive system related antigen DNA SEQ ID NO: 8207.

KW Human; reproductive system related antigen; reproductive system disorder;
cancer; gene therapy; ds.

XX Homo sapiens.

PN WO200155320-A2.

PD 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US01339.

XX 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 11-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234224.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.

PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249246.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Barash SC, Ruben SM;
XX WPI: 2001-465570/50.
DR
XX
PT Isolated nucleic acid molecule encoding a reproductive system antigen
XX is used in preventing, treating or ameliorating a medical condition -
PS
XX Disclosure: SEQ ID NO 8207; 1297PP + Sequence Listing; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention.
XX
SQ Sequence 24028 BP; 5272 A; 6390 C; 6884 G; 5482 T; 0 other;
Query Match 7.2%; Score 20; DB 22; Length 24028;
Best Local Similarity 100.0%; Pred. NO. 1.2;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 255 TTCGGAGCGCCGAGCAGAGA 274
DB 21264 TTCGGAGCGCCGAGCAGAGA 21283
RESULT 14
AAL05619
ID AAL05619 standard; DNA; 25574 BP.
XX
XX AAL05619;
AC
XX
DT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen DNA SEQ ID NO: 8307.

XX
KW Human; reproductive system related antigen; reproductive system disorder;
KW cancer; gene therapy; ds.
XX
XX Homo sapiens.
OS
XX
XX WO200155320-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01339.
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214886.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
XX 07-JUL-2000; 2000US-0216880.
XX 11-JUL-2000; 2000US-0217487.
XX 11-JUL-2000; 2000US-0217496.
XX 14-JUL-2000; 2000US-0218290.
XX 26-JUL-2000; 2000US-0220963.
XX 26-JUL-2000; 2000US-0220964.
XX 14-AUG-2000; 2000US-0224518.
XX 14-AUG-2000; 2000US-0224519.
XX 14-AUG-2000; 2000US-0225213.
XX 14-AUG-2000; 2000US-0225214.
XX 14-AUG-2000; 2000US-0225266.
XX 14-AUG-2000; 2000US-0225267.
XX 14-AUG-2000; 2000US-0225268.
XX 14-AUG-2000; 2000US-0225270.
XX 14-AUG-2000; 2000US-0225447.
XX 14-AUG-2000; 2000US-0225757.
XX 14-AUG-2000; 2000US-0225758.
XX 14-AUG-2000; 2000US-0225759.
XX 18-AUG-2000; 2000US-0226279.
XX 22-AUG-2000; 2000US-0226681.
XX 22-AUG-2000; 2000US-0226868.
XX 22-AUG-2000; 2000US-0227182.
XX 23-AUG-2000; 2000US-0227009.
XX 30-AUG-2000; 2000US-0228924.
XX 01-SEP-2000; 2000US-0229287.
XX 01-SEP-2000; 2000US-0229343.
XX 01-SEP-2000; 2000US-0229344.
XX 01-SEP-2000; 2000US-0229345.
XX 05-SEP-2000; 2000US-0229509.
XX 05-SEP-2000; 2000US-0229513.
XX 06-SEP-2000; 2000US-0230437.
XX 06-SEP-2000; 2000US-0230438.
XX 08-SEP-2000; 2000US-0231242.
XX 08-SEP-2000; 2000US-0231243.
XX 08-SEP-2000; 2000US-0231244.
XX 08-SEP-2000; 2000US-0231413.
XX 08-SEP-2000; 2000US-0231414.
XX 08-SEP-2000; 2000US-0232080.
XX 08-SEP-2000; 2000US-0232081.
XX 12-SEP-2000; 2000US-0231968.
XX 14-SEP-2000; 2000US-0233397.
XX 14-SEP-2000; 2000US-0233398.
XX 14-SEP-2000; 2000US-0233399.
XX 14-SEP-2000; 2000US-0233400.
XX 14-SEP-2000; 2000US-0233401.
XX 14-SEP-2000; 2000US-0233063.
XX 14-SEP-2000; 2000US-0233064.
XX 14-SEP-2000; 2000US-0233065.
XX 21-SEP-2000; 2000US-0234223.

21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239835.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249246.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249399.
PR 17-NOV-2000; 2000US-0249399.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251858.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.

PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM,
XX
DR WPI; 2001-465570/50.
XX
PT Isolated nucleic acid molecule encoding a reproductive system antigen
is used in preventing, treating or ameliorating a medical condition -
XX
PS Disclosure; SEQ ID NO 8307; 1297pp + Sequence Listing; English.
XX
CC The present invention provides the protein and coding sequences of a
number of human reproductive system related antigens. These can be used
in the prevention and treatment of reproductive system disorders.
CC Including cancer. The present sequence is a genomic sequence encoding a
protein of the invention.
CC
XX
SQ Sequence 25574 BP; 7154 A; 6018 C; 6058 G; 6344 T; 0 other;

Query Match 7.2%; Score 20; DB 22; Length 25574;
Best Local Similarity 100.0%; Pred. No. 1.2; Mismatches 0; Gaps 0;
Matches 20; Conservative 0; Indels 0;

OY 227 GTGATGCTCAGCCTGTAAT 246
DB 12248 GTGATGCTCAGCCTGTAAT 12267
|||||

RESULT 15
AAK79671
ID AAK79671 standard; DNA; 25574 BP.
XX
AC AAK79671;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human Immune/haematopoietic antigen genomic sequence SEQ ID NO:34483.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN MO200157182-A2.
PD 09-AUG-2001.
XX
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 23-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234223.
PR 25-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
XX

PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249267.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
PI WPI, 2001-483426/52.
XX
XX
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
XX
PS Disclosure: SEQ ID NO 34483; 3071bp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX

SQ Sequence 25574 BP; 7154 A; 6018 C; 6058 G; 6344 T; 0 other;

Query Match 7.2%; Score 20; DB 22; Length 25574;

Best Local Similarity 100.0%; Pred. No. 1.2;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 227 GTGATGGCTCAGCCTGTAAT 246

|||||

DB 12248 GTGATGGCTCAGCCTGTAAT 12267

Search completed: October 8, 2002, 13:22:25
Job time : 232 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 8, 2002, 12:25:08 : Search time 1810 Seconds
(Without alignments)
3191.008 Million cell updates/sec

Title: US-09-065-672-5_COPY_1_276

Perfect score: 276
Sequence: 1 CTRAGGCGTGCANACAGAGC.....CGGAGGCCGAGCAGGAG 276

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapept 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:*
1: gb_da:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
------------	-------	-------	-------	--------	----	----	-------------

1	225	81.5	381	9	AF331165	AF331165 Homo sapi
2	135	48.9	85424	2	AC091179	AC091179 Homo sapi
3	135	48.9	192077	2	AC068531	AC068531 Homo sapi
4	23	8.3	149948	2	AC068852	AC068852 Homo sapi
5	23	8.3	163204	2	AL356749	AL356749 Homo sapi
6	23	8.3	191764	2	AL358859	AL358859 Homo sapi
7	23	8.3	207486	2	AL356583	AL356583 Homo sapi
8	22	8.0	39631	9	AC004030	AC004030 Homo sapi
9	22	8.0	63187	9	AL355477	AL355477 Homo sapi
10	22	8.0	115710	2	AL138015	AL138015 Homo sapi
11	22	8.0	117096	2	AC008468	AC008468 Homo sapi
12	22	8.0	142094	2	HSJ543J19	AL109940 Homo sapi
13	22	8.0	149138	2	AC026936	AC026936 Homo sapi
14	22	8.0	154312	2	AC011492	AC011492 Homo sapi
15	22	8.0	157970	2	AC005609	AC005609 Homo sapi
16	22	8.0	181808	2	AL606465	AL606465 Homo sapi
17	22	8.0	189409	2	AL591122	AL591122 Homo sapi
18	22	8.0	206442	2	AL159978	AL159978 Homo sapi
19	22	8.0	218074	9	HUAC002044	AC002044 Homo sapi
20	22	8.0	230827	9	AC010102	AC010102 Homo sapi
21	22	8.0	337275	2	AL358788	AL358788 Homo sapi
22	21	7.6	2536	9	AK055900	AK055900 Homo sapi
23	21	7.6	13562	9	HSU70065	U70065 Homo sapi
24	21	7.6	37391	9	AC005779	AC005779 Homo sapi
25	21	7.6	41006	9	AC007201	AC007201 Homo sapi
26	21	7.6	51677	9	AL133288	AL133288 Homo sapi
27	21	7.6	52875	2	AC078968	AC078968 Homo sapi
28	21	7.6	64299	2	AC105084	AC105084 Homo sapi
29	21	7.6	68275	9	AC073335	AC073335 Homo sapi
30	21	7.6	68692	9	HSJ468N4	AL12152 Homo sapi
31	21	7.6	69270	2	AC103817	AC103817 Homo sapi
32	21	7.6	89448	9	AC005067	AC005067 Homo sapi
33	21	7.6	89468	2	AL365190	AL365190 Homo sapi
34	21	7.6	98878	2	AC087108	AC087108 Homo sapi
35	21	7.6	102892	2	AL359813	AL359813 Homo sapi
36	21	7.6	128811	2	HSJ764D2	AL122018 Homo sapi
37	21	7.6	135584	2	AC020749	AC020749 Homo sapi
38	21	7.6	140092	2	AL162615	AL162615 Homo sapi
39	21	7.6	144457	2	AC023268	AC023268 Homo sapi
40	21	7.6	150193	2	AC009641	AC009641 Homo sapi
41	21	7.6	150351	2	AP002505	AP002505 Homo sapi
42	21	7.6	152448	2	AC021391	AC021391 Homo sapi
43	21	7.6	159791	9	AC007179	AC007179 Homo sapi
44	21	7.6	160069	2	AC023329	AC023329 Homo sapi
45	21	7.6	161938	9	AL360227	AL360227 Homo sapi

ALIGNMENTS

RESULT 1
AF331165 LOCUS Homo sapiens PRAC mRNA, complete cds.
DEFINITION AF331165
ACCESSION AF331165
VERSION AF331165.1 GI:13774329
KEYWORDS

SOURCE

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 381)
AUTHORS Liu,X.F., Olsson,P., Wolfgang,C.D., Bera,T.K., Duray,P., Lee,B. and Pastan,I.

TITLE

PRAC: A novel small nuclear protein that is specifically expressed in human prostate and colon
JOURNAL Prostata 47 (2), 125-131 (2001)

MEDLINE

REFERENCE

AUTHORS

TITLE

2 (bases 1 to 381)
Liu,X.F., Olsson,P., Wolfgang,C.D., Bera,T., Duray,P., Lee,B. and Pastan,I.
Direct Submission
Submitted (20-DEC-2000) Lab Mol Biol, NCI/NIH, 37 Convent Dr. 4B20,

FEATURES
 source
 Bethesda, MD 20892, USA
 Location/Qualifiers
 1.381
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="17"
 /map="17q21"
 /tissue_type="prostate"
 129.302
 /codon_start=1
 /product="PRAC"
 /protein_id="AAK38837.1"
 /db_xref="GI:13774330"
 /translation="MLCAHFSDDQPAHLTTSKSAFLSNKSTSLKHLGELTRSDGSAC
 NSGISGGRKRIP"
 CDS
 BASE COUNT
 100 a 94 c 93 g 94 t
 ORIGIN
 Query Match 81.5%; Score 225; DB 9; Length 381;
 Best Local Similarity 99.6%; Pred. No. 1.7e-115;
 Matches 275; Conservative 0; Mismatches 1; Indels 0; Gaps 0:
 1 CTAAGGCGTCAACAGACGGCCACTGGAGGCTGAACCTTTAGGCGATGCTGCTTG 60
 18 CTAAGGCGTCAACAGACGGCCACTGGAGGCTGAACCTTTAGGCGATGCTGCTTG 77
 61 CAAAGTCAGCAGCAGCTGATTCGTGTCCTCCACTTTGCAGAGAGAACAGCAGATGTTGTC 120
 78 CAAAGTCAGCAGCAGCTGATTCGTGTCCTCCACTTTGCAGAGAGAACAGCAGATGTTGTC 137
 121 GCCCATTTTCGATCAGACGACCGCCCATTTTACTACTCTCCAAAGAGTCTTTCTCTCT 180
 138 GCCCATTTTCGATCAGACGACCGCCCATTTTACTACTCTCCAAAGAGTCTTTCTCTCT 197
 181 AATTAAGAAACATCTACTTTGAACATCTACTGGCGGAGACAGCAGATGGCTCAGCC 240
 198 AATTAAGAAACATCTACTTTGAACATCTACTGGCGGAGACAGCAGATGGCTCAGCC 257
 241 TGTAAATTCGTGAATTTTCGGAGGCGCAGCAGAGAG 276
 258 TGTAAATTCGTGAATTTTCGGAGGCGCAGCAGAGAG 293
 RESULT 2
 AC091179 85424 bp DNA linear PRI 22-NOV-2001
 LOCUS Homo sapiens chromosome 17, clone CTD-2377D24, complete sequence.
 DEFINITION AC091179
 ACCESSION AC091179.5 GI:17047193
 VERSION AC091179.5 GI:17047193
 KEYWORDS
 HTG.
 SOURCE
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 85424)
 2 (bases 1 to 85424)
 Homo sapiens chromosome 17, clone CTD-2377D24
 Unpublished
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
 Batra,N., Bastien,V., Boguslavsky,L., Boukhalter,B., Brown,A.,
 Camarata,J., Campolano,A., Chang,J., Choepel,Y., Colangelo,M.,
 Collins,S., Collymore,A., Cooke,P., Deatellano,K., Dewar,K.,
 Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., Fitzhugh,W., Gage,D.,
 Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
 Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W.,
 Iliev,I., Johnson,R., Jones,C., Karatas,A., Laroque,K.,
 Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
 Maclean,C., Macdonald,P., Margulis,N., Matthews,C., McCarthy,M.,
 McEwan,P., McKernan,K., McPheters,R., Meldrum,J., Menus,L.,
 Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,
 Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
 Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
 Rosetti,M., Roy,A., Santos,R., Schauer,S., Schnipack,R., Seaman,S.,
 Severy,P., Sougnuez,C., Spencer,B., Stange-Thomann,N.,
 Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
 Theodore,J., Travers,M., Travis,N., Trifillo,J., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (03-APR-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 85424)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
 Anderson,S., Batra,N., Bastien,V., Boguslavsky,L., Boukhalter,B.,
 Brown,A., Camarata,J., Campolano,A., Chang,J., Chazaro,B.,
 Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,A.,
 Cooke,P., Deatellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
 Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
 Ginde,S., Gind,S., Goyette,M., Graham,L., Grand-Pierre,N.,
 Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
 Jones,C., Kamat,A., Karatas,A., Kells,C., Laroque,K.,
 Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
 Maclean,C., Macdonald,P., Major,J., Margulis,N., Matthews,C.,
 McCarthy,M., McEwan,P., McKernan,K., McPheters,R., Meldrum,J.,
 Menus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
 Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
 Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
 Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schnipack,R.,
 Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
 Topham,K., Travers,M., Trivis,N., Trifillo,J., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Nov 22, 2001 this sequence version replaced gi:16905296.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RW/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: IJ3056
 Center clone name: 2377_D_24

 FEATURES
 source
 Location/Qualifiers
 1.85424
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="17"
 /map="17"
 /clone="CTD-2377D24"
 /clone_lib="CIRD Human BAC"
 complement(5..84)
 /rpt_family="MIR"
 593..636
 repeat_region
 /rpt_family="(TTCC)n"
 complement(1106..1234)
 /rpt_family="FLAM_C"
 complement(3707..3990)
 /rpt_family="AluJo"
 5533..5899
 repeat_region
 /rpt_family="L1MC4a"
 complement(6039..6185)
 /rpt_family="FLAM_C"
 8157..8197
 repeat_region
 /rpt_family="A-rich"
 8219..8266
 repeat_region
 /rpt_family="(CA)n"

```

repeat_region      10354. .10416
/rpt_family="TA)n"
repeat_region      11022. .11305
/rpt_family="Alusx"
repeat_region      complement(11698. .11814)
/rpt_family="MIR"
repeat_region      12352. .12583
/rpt_family="L1MC5"
repeat_region      12598. .12719
/rpt_family="FLAM_C"
repeat_region      12725. .12988
/rpt_family="Alu"
repeat_region      13033. .13059
/rpt_family="Alu"
repeat_region      13117. .13266
/rpt_family="Alu"
repeat_region      13284. .13317
/rpt_family="CAAAA)n"
repeat_region      13331. .13456
/rpt_family="L1MC5"
repeat_region      13458. .13757
/rpt_family="Alusg"
repeat_region      complement(14208. .14380)
/rpt_family="Alusx"
repeat_region      14381. .14404
/rpt_family="(TTTA)n"
repeat_region      complement(14405. .14524)
/rpt_family="Alusx"
repeat_region      14686. .15016
/rpt_family="Alu"
repeat_region      complement(15137. .15242)
/rpt_family="Alu"
repeat_region      complement(15262. .15353)
/rpt_family="Alu/FLAM"
repeat_region      15354. .15465
/rpt_family="MIR"
repeat_region      17066. .17183
/rpt_family="CT-rich"
repeat_region      complement(17200. .17513)
/rpt_family="Alusx"
repeat_region      complement(17747. .18048)
/rpt_family="Alu"
repeat_region      complement(19599. .19727)
/rpt_family="MIR3"
repeat_region      20715. .20952
/rpt_family="HAL1"
repeat_region      20970. .21087
/rpt_family="GA-rich"
repeat_region      21461. .21484
/rpt_family="GC-rich"
repeat_region      complement(22025. .22109)
/rpt_family="MIR"
repeat_region      22229. .22541
/rpt_family="Alusx"
repeat_region      22812. .22840
/rpt_family="AT-rich"
repeat_region      complement(22841. .23291)
/rpt_family="L1ME3"
repeat_region      complement(23361. .23683)
/rpt_family="Alusx"
repeat_region      complement(23684. .23770)
/rpt_family="FLAM_A"
repeat_region      complement(23775. .23919)
/rpt_family="L1ME3"
repeat_region      25706. .25851
/rpt_family="(GA)n"
repeat_region      27240. .27270
/rpt_family="(TG)n"
repeat_region      32764. .32804
/rpt_family="AT-rich"
repeat_region      32840. .32954
/rpt_family="FLAM_C"
repeat_region      33081. .33358
/rpt_family="Alusx"
repeat_region      33363. .33383
/rpt_family="(TA)n"
repeat_region      35452. .35518
/rpt_family="(TG)n"
repeat_region      complement(36800. .36936)
/rpt_family="L1ME4A"
repeat_region      36952. .37269
/rpt_family="Alu"
repeat_region      37291. .37347
/rpt_family="Alu"
repeat_region      37387. .37425
/rpt_family="AT-rich"
repeat_region      complement(37514. .37642)
/rpt_family="L1ME4A"
repeat_region      37810. .37905
/rpt_family="FRAM"
repeat_region      complement(37906. .38216)
/rpt_family="Alu"
repeat_region      38256. .38386
/rpt_family="L1ME3"
repeat_region      38814. .38841
/rpt_family="AT-rich"
repeat_region      complement(39326. .39493)
/rpt_family="L1ME3"
repeat_region      complement(39494. .39791)

Query Match      48.9%: Score 135; DB 9; Length 85424;
Best Local Similarity 99.5%: Pred. No. 4,5e-65;
Matches 185; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy      1 CTAAGGCGTGAACAGAGCGGCACTGGAGGCTGAACCTTTAGCCGATGCTG 60
Db      32202 CTAAGGCGTGAACAGAGCGGCACTGGAGGCTGAACCTTTAGCCGATGCTG 32261

Oy      61 CAAGGTCAAGCAAGCTGATTCGTGTCGCCACCTTTGCAGAGAGACGATGTTGTC 120
Db      32262 CAAGGTCAAGCAAGCTGATTCGTGTCGCCACCTTTGCAGAGAGACGATGTTGTC 32321

Oy      121 GCCCATTTCTCAGATCAAGAGCGGCCCATCTTACTACTCCAGAGTCTTTCTCT 180
Db      32322 GCCCATTTCTCAGATCAAGAGCGGCCCATCTTACTACTCCAGAGTCTTTCTCT 32381

Oy      181 AATAAG 186
Db      32382 AATAAG 32387

RESULT 3
AC068531/c      192077 bp      DNA      linear      HTG 29-NOV-2001
LOCUS      Homo sapiens chromosome 17 clone RP11-463M16 map 17, WORKING DRAFT
DEFINITION      SEQUENCE, 11 unordered pieces.
ACCESSION      AC068531
VERSION      AC068531.4 GI:16506910
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULFILLTOP.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 192077)
AUTHORS      Biren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Boguslavsky,L., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Campoliano,A., Castle,A., Choquel,Y., Colangelo,M., Collins,S.,
Collins,A., Cooke,P., Dearlano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heatford,A., Horton,L.,
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 192077)
AUTHORS      Biren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Boguslavsky,L., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Campoliano,A., Castle,A., Choquel,Y., Colangelo,M., Collins,S.,
Collins,A., Cooke,P., Dearlano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heatford,A., Horton,L.,

```

Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Lacroque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G., Locke, K., MacDonald, P., Marquis, N., McCarthy, M., McEwan, P., McGuirk, A., McKernan, K., McNeesters, R., Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mlepea, Y., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Olivari, T.M., Oliver, J., Peterson, K., Pierre, N., Pisanli, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Testa, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

FEATURES

Submitted (03-MAY-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Oct 28, 2001 this sequence version replaced gi:13959304.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: MIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: 463_M_16

Center clone name: 19798

Summary Statistics

Sequencing vector: M13; M77815; 35% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 188372 bases at least Q40

Consensus quality: 189604 bases at least Q30

Consensus quality: 190345 bases at least Q20

Insert size: 163000; agarose-fp

Insert size: 191077; sum-of-contigs

Quality coverage: 12.9 in Q20 bases; agarose-fp

Quality coverage: 11.0 in Q20.

NOTE: This is a 'working draft' sequence. It currently consists of 11 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 27168: contig of 27168 bp in length
27169 27268: gap of 100 bp
27269 27776: contig of 508 bp in length
27777 27876: gap of 100 bp
27877 31207: contig of 3331 bp in length
31208 31307: gap of 100 bp
31308 35824: contig of 4517 bp in length
35825 35924: gap of 100 bp
35925 86448: contig of 50524 bp in length
86449 86548: gap of 100 bp
86549 94772: contig of 8224 bp in length
94773 94872: gap of 100 bp
94873 109036: contig of 14164 bp in length
109037 109136: gap of 100 bp
109137 133011: contig of 23875 bp in length
133012 133111: gap of 100 bp
133112 155597: contig of 22486 bp in length
155598 155697: gap of 100 bp
155698 185359: contig of 29662 bp in length
185360 185458: gap of 100 bp
185459 192077: contig of 6618 bp in length.

Location/Qualifiers

1. 192077
/organism="Homo sapiens"
/db_xref="taxon:9606"

/chromosome="17"
/map="17"
/clone="RP11-463M16"
/clone_lib="RPC1-11 Human Male BAC"
1. 27168
/note="assembly-fragment
clone_end:SP6
vector_side:left"
27269..27776
/note="assembly-fragment"
27877..31207
/note="assembly-fragment"
31308..35824
/note="assembly-fragment"
35925..86448
/note="assembly-fragment"
86549..94772
/note="assembly-fragment"
94873..109036
/note="assembly-fragment"
109137..133011
/note="assembly-fragment"
133112..155597
/note="assembly-fragment"
155698..185359
/note="assembly-fragment"
185460..192077
/note="assembly-fragment
clone_end:T7
vector_side:right"
BASE COUNT 49460 a 44914 c 44716 g 51987 t 1000 others
ORIGIN

Query Match 48.9%; Score 135; DB 2; Length 192077;
Best local Similarity 99.5%; Pred. No. 4.1e-65;
Matches 185; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTAAAGCGTGCAGAAACAGAGCGCCACTGGAGGCTGAACCTTTAGCCGATGCTTGG 60
DB 17711 CTAAAGCGTGCAGAAACAGAGCGCCACTGGAGGCTGAACCTTTAGCCGATGCTTGG 17652
QY 61 CAAGGTCAGGCAAGCTGGATTGTCGCCACCTTGGCAGAGAGCAGAGATGTTGTC 120
DB 17651 CAAGGTCAGGCAAGCTGGATTGTCGCCACCTTGGCAGAGAGCAGAGATGTTGTC 17592
QY 121 GCCCATTTCTCAGATCAGACGCGCCACTTACTACTCTCAAGAGTCTTCTCTCT 180
DB 17591 GCCCATTTCTCAGATCAGACGCGCCACTTACTACTCTCAAGAGTCTTCTCTCT 17532
QY 181 AATAAG 186
DB 17531 AATAAG 17526

RESULT 4
LOCUS Homo sapiens clone RP11-371E1, WORKING DRAFT SEQUENCE, 14 unordered
AC068852
DEFINITION Homo sapiens clone RP11-371E1, WORKING DRAFT SEQUENCE, 14 unordered
pieces.
AC068852
VERSION AC068852.2 GI:8317993
KEYWORDS HTGS, HTGS_PHASE1, HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 149948)
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE Homo sapiens, clone RP11-371E1
JOURNAL Unpublished
2 (bases 1 to 149948)
REFERENCE Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bede, F.,

TITLE
JOURNAL
COMMENT

Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G.,
Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
Collamore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,
Dodde, S., Domino, M., Doyle, M., Ferreltra, P., Fitzhugh, W., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L.,
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., Larocque, K., Lamazares, R., Landers, T., Lehoczek, J.,
Levine, R., Lieu, G., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheters, R.,
Meldrum, J., Menus, L., Mihova, T., Miranda, C., Mienga, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Olyar, T.M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Strange, Rhoman, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Testaye, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Submitted (10-MAY-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 7, 2000 this sequence version replaced g1:7767857.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center project name: L8348

Center clone name: 371_E1

----- Summary Statistics

Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 141064 bases at least Q40
Consensus quality: 145546 bases at least Q30
Consensus quality: 147492 bases at least Q20
Insert size: 151000; agarose-fp
Insert size: 148648; sum-of-ctrls
Quality coverage: 4.6 in Q20 bases; agarose-fp
Quality coverage: 4.7 in Q20 bases; sum-of-ctrls

* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

* 1 1333: contig of 1333 bp in length
* 1334 1433: gap of 100 bp
* 1434 2803: contig of 1370 bp in length
* 2804 2903: gap of 100 bp
* 2904 4034: contig of 1131 bp in length
* 4035 4134: gap of 100 bp
* 4135 6009: contig of 1875 bp in length
* 6010 6109: gap of 100 bp
* 6110 8505: contig of 2396 bp in length
* 8506 8605: gap of 100 bp
* 8606 10785: contig of 2180 bp in length
* 10786 10885: gap of 100 bp
* 10886 16511: contig of 5766 bp in length
* 16512 16751: gap of 100 bp
* 16752 23392: contig of 6641 bp in length
* 23393 23432: gap of 100 bp
* 23433 30190: contig of 6698 bp in length
* 30191 30290: gap of 100 bp
* 30291 42641: contig of 12351 bp in length

```

```

* 42642 42741: gap of 100 bp
* 42742 56837: contig of 14096 bp in length
* 56838 56937: gap of 100 bp
* 56938 79527: contig of 22550 bp in length
* 79528 79627: gap of 100 bp
* 79628 107159: contig of 27532 bp in length
* 107160 107259: gap of 100 bp
* 107260 149948: contig of 42689 bp in length.
Location/Qualifiers
1. 149948
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-371E1"
/clone_lib="RP11-371 Human Male BAC"
1. 1333
/feature="assembly_fragment"
1434. 2803
/feature="assembly_fragment"
2904. 4034
/feature="assembly_fragment"
4135. 6009
/feature="assembly_fragment"
6110. 8505
/feature="assembly_fragment"
8606. 10785
/feature="assembly_fragment"
10886. 16511
/feature="assembly_fragment"
16752. 23392
/feature="assembly_fragment"
23493. 30190
/feature="assembly_fragment"
clone_end:r7
vector_side:right"
30291. 42641
/feature="assembly_fragment"
42742. 56837
/feature="assembly_fragment"
56938. 79527
/feature="assembly_fragment"
79628. 107159
/feature="assembly_fragment"
107260. 149948
/feature="assembly_fragment"
clone_end:SP6
vector_side:right"

```

BASE COUNT 43445 a 28809 c 29291 g 47098 t 1305 others
ORIGIN

Query Match 8.3%; Score 23; DB 2; Length 149948;
Best Local Similarity 100.0%; Pred. No. 0.058;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 251 GAATTCGAGCGAGCGAGCAGG 273

Db 134092 GAATTCGAGCGAGCGAGCAGG 134114

```

RESULT 5 163204 bp DNA linear HTG 10-JUN-2001
AL356749 Homo sapiens chromosome 1 clone RP11-831112, *** SEQUENCING IN
LOCUS AL356749
DEFINITION PROGRESS *** 12 unordered pieces.
ACCESSION AL356749
VERSION AL356749.11 GI:10039925
KEYWORDS HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 163204)
AUTHORS McInay, K.
TITLE Direct Submission

```

JOURNAL

Submitted (10-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
On Sep 9, 2000 this sequence version replaced gi:9930936.
----- Genome Center

COMMENT

Center: Sanger Centre

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: humquery@sanger.ac.uk

----- Project Information

Center project name: ba831112

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Sequencing vector: plasmid; L08752; 100% of reads

Chemistry: Dye-terminator ET-amersham; 24% of reads Chemistry:

Dye-terminator Big Dye; 75% of reads

Consensus quality: 159001 bases at least Q40

Consensus quality: 160363 bases at least Q30

Consensus quality: 161238 bases at least Q20

Insert size: 162104; sum-of-contigs

Insert size: 169741; 7.0% error; agarose-fp

Quality coverage: 6.87x in Q20 bases; sum-of-contigs Quality

coverage: 6.70x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 2484: contig of 2484 bp in length
* 2485 2584: gap of 100 bp
* 2585 4801: contig of 2217 bp in length
* 4802 4901: gap of 100 bp
* 4902 7732: contig of 2831 bp in length
* 7733 7832: gap of 100 bp
* 7833 13979: contig of 6147 bp in length
* 13980 14079: gap of 100 bp
* 14080 27378: contig of 13299 bp in length
* 27379 27478: gap of 100 bp
* 27479 39855: contig of 12377 bp in length
* 39856 39955: gap of 100 bp
* 39956 42487: contig of 2532 bp in length
* 42488 42587: gap of 100 bp
* 42588 48516: contig of 5929 bp in length
* 48517 48616: gap of 100 bp
* 48617 56232: contig of 7616 bp in length
* 56233 56332: gap of 100 bp
* 56333 148274: contig of 91942 bp in length
* 148275 148374: gap of 100 bp
* 148375 151345: contig of 2971 bp in length
* 151346 151445: gap of 100 bp
* 151446 163204: contig of 11759 bp in length.

FEATURES

SOURCE

1. 163204

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="1"

/clone="RP11-831112"

/clone_11b="RPCT-11.3"

1. 2484

/note="assembly-fragment:02263"

fragment_chain:1"

2585. 4801

/note="assembly-fragment:00336"

fragment_chain:1"

4902. 7732

/note="assembly-fragment:02502"

fragment_chain:1"

7833. 13979

/note="assembly-fragment:01687"

misc.feature

fragment_chain:1"
14080. 27378
/note="assembly-fragment:01027"

misc.feature

fragment_chain:1"
27479. 39855
/note="assembly-fragment:00391"

misc.feature

fragment_chain:1"
39956. 42487
/note="assembly-fragment:02612"

misc.feature

fragment_chain:1"
42588. 48516
/note="assembly-fragment:02551"

misc.feature

fragment_chain:2"
48617. 56232
/note="assembly-fragment:01565"

misc.feature

fragment_chain:2"
56333. 148274
/note="assembly-fragment:00168"

misc.feature

148375. 151345
/note="assembly-fragment:00707"

BASE COUNT

50177 a 32704 c 30910 g 48308 t 1105 others

ORIGIN

Query Match

Best Local Similarity 8.3%; Score 23; DB 2; Length 163204;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 251 GAATTCGGAGCGCGAGCAGG 273

DB 106117 GAATTCGGAGCGCGAGCAGG 106139

RESULT 6

AL358859

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AL358859 191764 bp DNA linear HTG 30-JAN-2002
Homo sapiens chromosome 1 clone RP11-545G13, *** SEQUENCING IN
PROGRESS ***, 3 unordered pieces.
AL358859
AL358859.19 GI:18476590
HTG: HTGS_PHASE1; HTGS_ACTIVERFIN; HTGS_DRAFT; HTGS_FILLTOP.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Nickerson,T.
Direct Submission
Submitted (29-JUN-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk
On Feb 1, 2002 this sequence version replaced gi:18375799.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: ba545G13
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator ET-amersham; 1% of reads Chemistry:
Dye-terminator Big Dye; 98% of reads
Consensus quality: 191270 bases at least Q40
Consensus quality: 191380 bases at least Q30
Consensus quality: 191480 bases at least Q20
Insert size: 191564; sum-of-contigs
Insert size: 167427; 6.3% error; agarose-fp
Quality coverage: 7.87x in Q20 bases; sum-of-contigs Quality
coverage: 9.24x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES	
*	1 48175: config of 48175 bp in length
*	48176 48275: gap of 100 bp
*	48276 80779: config of 32504 bp in length
*	80780 80879: gap of 100 bp
*	80880 191764: config of 110885 bp in length
*	Location/Qualifiers

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-545G13"
/clone_1b="RPC1-11.2"
misc_feature
1..48175

```

```
misc_feature      fragment_chain:1"
                   48276. 80779
                   /note="assembly_fragment:01483
misc_feature      fragment_chain:1"
                   80880. 191764
```

BASE COUNT	a	c	g	t	
ORIGIN	62425	38822	35642	54675	200 others

	Query Match	8.3%	Score 23	DB 2	Length 191764
	Best Local Similarity	100.0%	Pred. No. 0.056		
	Matches 23	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	251 GAATTTGCGAGGCCGAGCCAGCAGG	273			
Db	176075 GAATTTGCGAGGCCGAGCCAGCAGG	176097			

RESULT 7				
AL356583				
LOCUS	207486 bp	DNA	linear	PRI 16-NOV-2001
DEFINITION	Human DNA sequence from clone RPL1-69023 on chromosome 1, complete			

ACCESSION	AL356583	
VERSION	AL356583.28	GI:16973837
KEYWORDS	HTG.	
SOURCE	human	

REFERENCE
AUTHORS
TITLE
JOURNAL

Eukaryota: Chordata: Craniata: Vertebrata: Euteleostomi:
Mammalia: Eutheria: Primates: Catarrhini: Homnidae: Homo.
1 (pages 1 to 207486)
Coville, G.
Direct Submission
Submitted (16-NOV-2001) Wellcome Trust Sanger Institute, Hinxton.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality > 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following

abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Emu, EMBL; Sw: SWISSPROT; Tr, TREMBL; Wp, WORMPEP; information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/MGP/Chrl> RP11-690C23 is from the library RPc1-11.3 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pBACE3.6 This sequence is the entire insert of clone RP11-690C23 The true right end of clone RP11-545613 is at 65479 in this sequence.

FEATURES
source

```
misc_feature 107841..108114
              /note="Single clone region. Sequence from clone PCR only."
misc_feature 107850
              /note="Single clone region. Sequence from clone PCR only."
misc_feature 122563..122601
```

BASE COUNT	63729	a	45864	c	42290	g	55603	t
ORIGIN								

Query Match	8.3%	Score 23	DB 9	Length 207486
Best Local Similarity	100.0%	Pred. No. 0.055		
Matches 23	Conservative 0	Mismatches 0	Indels 0	Gaps 0
OY	251	GAATTTCGGAGCGCCGAGCGCAG	273	
db	49791	GAATTTCGGAGCGCCGAGCGCAG	49813	

RESULT 8	
AC004030/c	
LOCUS	AC004030
DEFINITION	Homo sapiens DNA from chromosome 19, cosmid F71856, complete
	39631 bp
	DNA linear
	PRI 23-JAN-1998

ACCESSION	AC004030
VERSION	AC004030.1
KEYWORDS	HTG.
SOURCE	human.

REFERENCE
AUTHORS

Eukaryota; Metazoa; Chordates; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Plimates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 39631)
Iameddin, J. E., McCready, P. M., Skowronski, E., Adamson, A. W.,
Burkhardt-Schultz, K., Gordon, L., Kyle, A., Ramirez, M., Stilwagen, S.,
Garnes, J., Danganan, L., Poundstone, P., Christensen, M.

REFERENCE

Sequence analysis of a 3.5 Mb contig in 19p13.3 between CDC34 and D19S342

JOURNAL
REFERENCE
Unpublished
2 (bases 1 to 39631)

JOURNAL
Submitted (23-JAN-1998) Joint Genome Institute, Lawrence Livermore
National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
E-mail: jgibson@llnl.gov

```
source
1. .39631
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
```

```
/map="19p13.3 between CDC34 and D19S342"
/clone="F21856"
/cell_line="UV5HL9-5B"
/clone_lib="LL19NC02 F chromosome 19-specific cosmid
library"
/note="cosmid library constructed at LNL from flow-sorted
chromosomes from hybrid UV5HL9-5B, which carries
chromosome 19 as its only human chromosome"
326..601
/rpt_family="Alu"
repeat_region
misc_feature
687..837
/note="DPS similarity to g111665807|gnl|PID|d1014090
(D87460) KIAA0270 [Homo sapiens] (49..98); 100%
identity. Predicted exon, program: grail2exons_human_1.3,
frame: 0, quality: excellent, score: 100.000"
join(687..837,3765..3786,5611..5670,9944..10075,
15877..16406)
/note="human protein of unknown function, partial coding
sequence"
/codon_start=2
/product="KIAA0270"
/protein_id="AAB97619.1"
/db_xref="gi:2804591"
/translation="LEKELEVLREGDSAPATAKENAAPSPVAPAPAPARERKTEV
VMNSQPTVGTGPKDKRVNTPLRTVDSAPMKAAAYSVETYEKRYGGERVLSST
LLRPRLPLGIKIVIEDERKVVHADVGTAKENGTHPLSSSEVDLHKADVTLSEGST
AGAAETRAVEGAARTPTSPREITGVQAQPGATIQGQEPPTMTFMGQNV
EDEATKRVLGIDDTITAEVLVIEDAAPKEPAPNGSAAPPTAASRENOAQPEA
TTSPODDMMKHKKCCSIM"
complement(1279..1531)
/rpt_family="Alu"
repeat_region
complement(1618..1780)
/rpt_family="Alu"
repeat_region
2005..2295
/rpt_family="Alu"
complement(2522..2808)
/rpt_family="Alu"
repeat_region
3308..3567
/rpt_family="Alu"
misc_feature
3765..3786
/note="DPS similarity to g111665807|gnl|PID|d1014090
(D87460) KIAA0270 [Homo sapiens] (99..106); 100%
identity."
3885..4415
/rpt_family="Alu"
repeat_region
misc_feature
5611..5670
/note="DPS similarity to g111665807|gnl|PID|d1014090
(D87460) KIAA0270 [Homo sapiens] (107..126); 100%
identity. Predicted exon, program: grail2exons_human_1.3,
frame: 2, quality: excellent, score: 89.000"
6360..6651
/rpt_family="Alu"
repeat_region
misc_feature
7477..7535
/note="Predicted exon, program: grail2exons_human_1.3,
frame: 0, quality: good, score: 70.000"
7710..8129
/rpt_family="Alu"
repeat_region
9003..9571
/rpt_family="Alu"
repeat_region
9944..10075
/rpt_family="Alu"
misc_feature
/note="DPS similarity to g111665807|gnl|PID|d1014090
(D87460) KIAA0270 [Homo sapiens] (127..170); 100%
identity. Predicted exon, program: grail2exons_human_1.3,
frame: 0, quality: excellent, score: 98.000"
10454..10701
/rpt_family="Alu"
repeat_region
complement(11471..11565)
/rpt_family="Alu"
repeat_region
11710..11753
/rpt_family="Alu"
repeat_region
11923..12211
/rpt_family="Alu"
repeat_region
```

```
misc_feature
complement(13647..13991)
/note="DPS similarity to AA340747 EST46017 Fetal kidney II
Homo sapiens cDNA 3' end similar to EST containing Alu
repeat. Score: 670 Identity: 340/345 (98%)."
13711..14006
/rpt_family="Alu"
repeat_region
14214..14468
/rpt_family="Alu"
repeat_region
14499..14777
/rpt_family="Alu"
repeat_region
15012..15294
/rpt_family="Alu"
repeat_region
15379..15657
/rpt_family="Alu"
repeat_region
15877..16403
/note="DPS similarity to g111665807|gnl|PID|d1014090
(D87460) KIAA0270 [Homo sapiens] (171..345); 100%
identity.-(15877..16380) predicted exon, program:
grail2exons_human_1.3
frame: 0, quality: excellent, score: 90.000"
16334..17921
/note="DPS similarity to overlapping ESTs:
(16334..16708) H4167 ym62f07.r1 Homo sapiens cDNA clone
163525 5'. Score: 724 Identity: 372/377
(98%)-(16334..16735) H4179 ym62h07.r1 Homo sapiens cDNA
clone 163549 5'. Score: 756 Identity: 397/406
(97%)-(16584..16949) D56226|HMA420E08B Human fetal brain
cDNA 5'-end GENE-420E08. Score: 699 Identity: 359/365
(98%)-(17162..16845) AA565621 nm8901.s1 NC1 CGAP_P4.1
Homo sapiens cDNA clone IMAGE:1062480. Score: 525
Identity: 301/319 (94%)-(17098..17588) AA233201
zr69a10.r1 Soares NHMPU S1 Homo sapiens cDNA clone 668634
5'. Score: 948 Identity: 488/489 (98%)-(17384..17921)
AA442399 zv70c01.r1 Soares total fetus NB2HF8 9w Homo
sapiens cDNA clone 758976 5'. Score: 1008 Identity:
529/539 (98%)-(17920..17412) N25536 yx76d01.s1 Homo
sapiens cDNA clone 267649 3'. Score: 882 Identity:
497/519 (95%)-(17921..17542) AA232968 zr69a10.s1 Soares
NHMPU S1 Homo sapiens cDNA clone 668634 3'. Score: 711
Identity: 376/389 (96%)-(17589..17544) H28126 y078c03.s1
Homo sapiens cDNA clone 184036 3'. Score: 687 Identity:
351/355 (98%)-(17921..17444) R60222 yhl3e04.s1 Homo
sapiens cDNA clone 42841 3'. Score: 745 Identity: 450/499
(90%)-(17921..17598) AA233859 zr6c01.s1 Soares NHMPU S1
Homo sapiens cDNA clone 666432 3'. Score: 636 Identity:
321/324 (99%)-and others . . .
18449..18737
/rpt_family="Alu"
repeat_region
19160..19436
/rpt_family="Alu"
repeat_region
complement(19791..20081)
/rpt_family="Alu"
repeat_region
21935..22125
/rpt_family="Alu"
repeat_region
22342..22853
/rpt_family="Alu"
misc_feature
/note="DPS similarity to AA577849 np24h02.s1 NCI CGAP-Gas1
Homo sapiens cDNA clone IMAGE:1084851. Score: 1002
Identity: 509/510 (99%)
complement(22982..23253)
/rpt_family="Alu"
repeat_region
complement(23377..23813)
/rpt_family="Alu"
repeat_region
23637..24227
/rpt_family="Alu"
repeat_region
25247..25537
/rpt_family="Alu"
repeat_region
join(26539..28318,29501..29631,31217..31255,33093..33182)
/note="hypothetical human protein of unknown function"
/product="F21856_2"
/protein_id="AAB97620.1"
/db_xref="gi:2804592"
/translation="MDRVTRYPILGIPQAHKRGTVLGDGDSITYTHLYCMGPEASGNG
```

misc.feature

0DEPQTWPTDHRAGQGVORQSVSYSHVATGYGQSPRGJHSENREDEGMVYRLCARD
 HOGPFWALPDEDEKEMKTYRLADADAPRLCDLEREMWAVIOCAVRSSTVAT
 LGCPDGDPRTPGPRSTPLEENVVDREOIDEFLAAROFLELEAGAPHSPPARG
 TPAGTPGASQAPKAFKPHLANGHVVPKQYGVREKNKAVATVMAVQVYDDP
 GSIAVSFSGTPKPTPIEREIRLAQEREADREORGJHQAATDHOELVETPRPLRLA
 SLTAPRRERGRPSLYVORDIYQTOREEDHRRGLHVGRASTPDWSEGPOLRLA
 LSSDSILSPADAADAPAREKRVNRIPPADYOPYLPSPGPQAFGAFGRPSL
 STAEKAASTPKATSPKRLSESSGKPLSTKQYAPKPPGPGQANRGVAVREYRLP
 LRFRAPDEPOAOVPHYWGMEVAGAPALRLQKSSDLLEERSVYLRDEGVLEER
 NALFEVFSPTPDNSQNSRSSQASGITGSYSVSESPRPHLSNVAWYEDV
 DSAPGQRKKEQWTAGINPDSGINSEVLAIKRVIRKNAAEWESRISVASEED"
 26539..28318
 /note="Overlapping exon prediction and ESTs matches:
 (26539..28560) predicted exon, program:
 grallexons human_1.3, frame: 0, quality: excellent,
 score: 93.000-Other overlapping matches:
 (26809..26520) AA213683 zq92h01.r1 Stratagene hmt neuron
 (#937233) Homo sapiens cDNA clone 649489 5',
 score: 432 identity: 268/302 (88%)~(26873..27134)
 AA552431 nk15c04.sl NCI.CCAP.C02 Homo sapiens cDNA clone
 IMAGE:1013574, score: 504 identity: 262/262
 (100%)~(27960..28318) AA160684 z072g01.r1 Stratagene
 pancreas (#937208) Homo sapiens cDNA clone 592464 5',
 (1..358): 99% identity~(28000..28318) AA179443 zpa5b09.r1

Query Match

Best Local Similarity 100.0%; Pred. No. 0.25; Length 39631;
 Matches 22: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 253 ATTTCGGAGCGCCGAGCAGCA 274
 |||

Db 11568 ATTTCGGAGCGCCGAGCAGCA 11547

RESULT 9
 AL355477
 LOCUS
 DEFINITION Human DNA sequence from clone RP11-113D13 on chromosome 1, complete
 sequence.
 ACCESSION AL355477
 VERSION
 KEYWORDS
 SOURCE HTG.
 ORGANISM human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 63187)
 Tracey, A.
 Direct Submission
 Submitted (16-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries: clonerequests@sanger.ac.uk
 humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
 On Nov 17, 2001 this sequence version replaced g1:16304508.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >= 30);
 an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. The following
 abbreviations are used to associate primary accession numbers given
 in the feature table with their source databases: Em., EMBL, SW.;
 SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP
 database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping

FEATURES

source

Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chrl>
 RP11-113D13 is from the library RPCI-11.1 constructed by the group
 of Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
 VECTOR: pBACe3.6
 IMPORTANT: This sequence is not the entire insert of clone
 RP11-113D13. It may be shorter because we sequence overlapping
 sections only once, except for a short overlap.
 The true left end of clone RP11-113D13 is at 1 in this sequence.
 The true left end of clone RP11-416A14 is at 61188 in this
 sequence. The true right end of clone RP11-781D11 is at 61193 in
 this sequence.

misc.feature
 BASE COUNT 15925 a 14194 c 15421 g 17647 t
 ORIGIN

Query Match

Best Local Similarity 100.0%; Pred. No. 0.24; Length 63187;
 Matches 22: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 253 ATTTCGGAGCGCCGAGCAGCA 274
 |||

Db 19512 ATTTCGGAGCGCCGAGCAGCA 19533

RESULT 10
 AL139015/c
 LOCUS
 DEFINITION Homo sapiens chromosome 1 clone RP4-648J17 map p34.1-34.3, ***
 SEQUENCING IN PROGRESS ***. 6 unordered pieces.
 ACCESSION AL139015
 VERSION
 KEYWORDS
 SOURCE HTG; HTGS_PHASE1; HTGS_CANCELLED.
 ORGANISM human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 115710)
 Plumb, B.
 Direct Submission
 Submitted (05-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 requests: clonerequests@sanger.ac.uk
 On Apr 9, 2001 this sequence version replaced g1:9796296.
 Genome Center
 Center: Sanger Centre
 Center code: SC
 Web site: <http://www.sanger.ac.uk>
 Contact: humquerry@sanger.ac.uk
 Project Information
 Center project name: d0648J17
 Summary Statistics
 Sequencing program: XGAP4; version 4.5
 Sequencing vector: pGAP4; version 1.08752; 100% of reads
 Chemistry: dye-terminator Big Dye; 100% of reads
 Consensus quality: 113662 bases at least Q40
 Consensus quality: 114216 bases at least Q30
 Consensus quality: 114601 bases at least Q20
 Insert size: 115210; sum-of-contigs
 Insert size: 132500; 15.0% error; agarose-ip
 Quality coverage: 6.13x in Q20 bases; sum-of-contigs quality
 coverage: 5.53x in Q20 bases; agarose-ip

* NOTE: This is a 'working draft' sequence. It currently

* consists of 6 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 36330: contig of 36330 bp in length
 * 36331 36430: gap of 100 bp
 * 36431 47455: contig of 11025 bp in length
 * 47456 47555: gap of 100 bp
 * 47556 52799: contig of 5244 bp in length
 * 52800 52899: gap of 100 bp
 * 52900 60268: contig of 7369 bp in length
 * 60269 60368: gap of 100 bp
 * 60369 88988: contig of 28620 bp in length
 * 88989 89088: gap of 100 bp
 * 89089 115710: contig of 26622 bp in length.

Location/Qualifiers
 1..115710
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /map="p34.1-34.3"
 /clone="RP4-648J17"
 /clone_lib="RP4-4"
 1..36330
 /note="assembly-fragment:01504
 fragment_chain:1"
 36431..47455
 /note="assembly-fragment:01208
 fragment_chain:1"
 47556..52799
 /note="assembly-fragment:00376
 fragment_chain:1"
 52900..60268
 /note="assembly-fragment:01176
 fragment_chain:1"
 60369..88988
 /note="assembly-fragment:00586
 fragment_chain:1"
 89089..115710
 /note="assembly-fragment:00793.0"

FEATURES

SOURCE

BASE COUNT 31095 a 27317 c 26251 g 30545 t 502 others
 ORIGIN

Query Match 8.0%; Score 22; DB 2; Length 115710;
 Best Local Similarity 100.0%; Pred. No. 0.22;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 253 ATTTCGGAGCGCCGAGCAGCA 274
 |||||
 Db 107684 ATTTCGGAGCGCCGAGCAGCA 107663

misc_feature
 /note="assembly-fragment:01176
 fragment_chain:1"
 52900..60268
 /note="assembly-fragment:01176
 fragment_chain:1"
 60369..88988
 /note="assembly-fragment:00586
 fragment_chain:1"
 89089..115710
 /note="assembly-fragment:00793.0"

misc_feature
 /note="assembly-fragment:01208
 fragment_chain:1"
 47556..52799
 /note="assembly-fragment:00376
 fragment_chain:1"

misc_feature
 /note="assembly-fragment:01176
 fragment_chain:1"
 52900..60268
 /note="assembly-fragment:01176
 fragment_chain:1"
 60369..88988
 /note="assembly-fragment:00586
 fragment_chain:1"

misc_feature
 /note="assembly-fragment:00586
 fragment_chain:1"
 89089..115710
 /note="assembly-fragment:00793.0"

RESULT 11
 AC008468/c AC008468 117096 bp DNA linear PRI 15-JUN-2000
 LOCUS Homo sapiens chromosome 5 clone CTC-365B8, complete sequence.
 DEFINITION AC008468
 ACCESSION AC008468
 VERSION AC008468.6 GI:8567760
 KEYWORDS HTC.
 ORGANISM human.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 117096)
 Direct Submission
 DOE Joint Genome Institute and Stanford Human Genome Center.
 JOURNAL
 TITLE Unpublished
 AUTHORS 2 (bases 1 to 117096)
 REFERENCE DOE Joint Genome Institute.
 TITLE Direct Submission

JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 3 (bases 1 to 117096)
 REFERENCE
 AUTHORS
 TITLE Direct Submission
 JOURNAL
 Submitted (15-JUN-2000) DOE Joint Genome Institute, 2800 Mitchell
 Drive, Walnut Creek, CA 94598, USA
 On Jun 15, 2000 this sequence version replaced gi:7711273.
 Draft Sequence Produced by DOE Joint Genome Institute
 www.jgi.doe.gov

COMMENT

Finishing completed at Stanford Human Genome Center
 www.shgc.stanford.edu
 Quality: Phrap quality >=40 99.8% of sequence;
 Estimated Total Number of Errors is 0.1.

FEATURES

SOURCE

BASE COUNT 38063 a 24433 c 22862 g 31738 t
 ORIGIN

Query Match 8.0%; Score 22; DB 9; Length 117096;
 Best Local Similarity 100.0%; Pred. No. 0.22;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 253 ATTTCGGAGCGCCGAGCAGCA 274
 |||||
 Db 77442 ATTTCGGAGCGCCGAGCAGCA 77421

misc_feature
 /note="assembly-fragment:01176
 fragment_chain:1"
 52900..60268
 /note="assembly-fragment:01176
 fragment_chain:1"
 60369..88988
 /note="assembly-fragment:00586
 fragment_chain:1"

RESULT 12
 HSJ543J19 142094 bp DNA linear PRI 26-FEB-2001
 LOCUS Human DNA sequence from clone RP4-543J19 on chromosome 20 contains
 DEFINITION (G protein, alpha stimulating activity polypeptide 1) including
 neuroendocrine secretory protein 55 (NEPS55), the CRTSA gene
 encoding cathepsin Z, the ATP5F gene encoding ATP synthase (H+
 transporting, mitochondrial F1 complex, epsilon subunit), the gene
 encoding protein HSPC130 (TH1 Drosophila homolog), the gene for
 tubulin beta 1 class VI (TUBB1), a gene encoding the GGT-107
 protein (LOC51012), four CpG islands, ESTs, STSs and GSSs, complete
 sequence.
 ACCESSION AL109840
 VERSION AL109840.24 GI:9369301
 KEYWORDS HTC; ATP5F; CRTSA; G protein; GNAS1; HSPC130; LOC51012; NEPS55;
 neuroendocrine secretory protein; TH1; TUBB1; tubulin.
 ORGANISM human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 142094)
 Direct Submission
 Moore, M.
 Submitted (26-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk
 On Jul 22, 2000 this sequence version replaced gi:9184434.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 The following abbreviations are used to associate primary accession
 numbers given in the feature table with their source/databases:
 Em.; EMBL; Sw.; SWISSPROT; Tr.; TrEMBL; Wp.; WORMPEP; Information
 on the WORMPEP database can be found at
 http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 20, constructed by the Sanger Centre Chromosome 20


```
repeat_region 22812..22929
/note="MIR repeat: matches 16. .139 of consensus"
repeat_region 23033..23076
/note="22 copies 2 mer tg 75% conserved"
repeat_region 23079..23116
/note="19 copies 2 mer gt 84% conserved"
repeat_region 23147..23279
/note="Aluub repeat: matches 7. .139 of consensus"
repeat_region 23288..23457
/note="MER30 repeat: matches 1. .196 of consensus"
repeat_region 24346..24656
/note="Aluub repeat: matches 1. .312 of consensus"
repeat_region 24798..24950
/note="MUR11 repeat: matches 254. .410 of consensus"
repeat_region 25020..25163

Query Match      8.0%: Score 22: DB 9: Length 142094:
Best Local Similarity 100.0%: Pred. No. 0.21:
Matches 22: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

QY 255 TTCGGAGCGCCGAGCAGAG 276
|||||
Db 96040 TTCGGAGCGCCGAGCAGAG 96061

RESULT 13
AC026936 149138 bp DNA linear HTG 27-APR-2000
LOCUS Homo sapiens chromosome 1 clone RP11-186C2 map 1, WORKING DRAFT
DEFINITION
AC026936 AC026936.2 GI:7652013
VERSION HTG: HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 149138)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 1, clone RP11-186C2
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 149138)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bedalov,F.,
Boguski,M., Boulikas,A., Brown,A., Burkett,G.,
Campbell,A., Castelle,A., Chao,J., Colangelo,M., Collins,S.,
Collumore,A., Cooke,P., Dearlano,K., Dewar,K., Diaz,J.S.,
Dodgson,S., Domingo,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardy,J., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heath,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Lander,T., Lechoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., MacDonald,P., Margolis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., Mcheters,R.,
Meldrum,J., Meneses,L., Mihova,T., Miranda,C., Mlenka,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollard,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Teschke,S., Theodore,J., Tirrell,A., Travers,M., Triggillo,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (25-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Apr 27, 2000 this sequence version replaced gi:7328801.
All repeats were identified using RepeatMasker:
Smt, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
```

```
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: 17617
Center clone name: 186.C-2
----- Summary Statistics
Sequencing vector: M13; W77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 132550 bases at least Q40
Consensus quality: 140551 bases at least Q30
Consensus quality: 144026 bases at least Q20
Insert size: 157000; agarose-1p
Insert size: 145938; sum-of-ctgigs
Quality coverage: 4.0 in Q20 bases; agarose-1p
Quality coverage: 4.3 in Q20 bases; sum-of-ctgigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 33 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 1364: contig of 1364 bp in length
* 1365 1464: gap of 100 bp
* 1465 2866: contig of 1462 bp in length
* 2867 2966: gap of 100 bp
* 2967 4325: contig of 1359 bp in length
* 4326 4425: gap of 100 bp
* 4426 5455: contig of 1030 bp in length
* 5456 5555: gap of 100 bp
* 5556 6761: contig of 1206 bp in length
* 6762 6861: gap of 100 bp
* 6862 8239: contig of 1438 bp in length
* 8300 8399: gap of 100 bp
* 8400 11139: contig of 2740 bp in length
* 11140 11239: gap of 100 bp
* 11240 13012: contig of 1773 bp in length
* 13013 13112: gap of 100 bp
* 13113 15320: contig of 2208 bp in length
* 15321 15420: gap of 100 bp
* 15421 16628: contig of 1408 bp in length
* 16629 16928: gap of 100 bp
* 16929 18769: contig of 1841 bp in length
* 18770 18869: gap of 100 bp
* 18870 20670: contig of 1801 bp in length
* 20671 20770: gap of 100 bp
* 20771 22998: contig of 2228 bp in length
* 22999 23098: gap of 100 bp
* 23099 23754: contig of 656 bp in length
* 23755 23854: gap of 100 bp
* 23855 26081: contig of 2227 bp in length
* 26082 26181: gap of 100 bp
* 26182 28619: contig of 2638 bp in length
* 28620 28919: gap of 100 bp
* 28920 30950: contig of 2031 bp in length
* 30951 31050: gap of 100 bp
* 31051 34127: contig of 3077 bp in length
* 34128 34227: gap of 100 bp
* 34228 37535: contig of 3308 bp in length
* 37536 37635: gap of 100 bp
* 37636 40364: contig of 2739 bp in length
* 40365 40464: gap of 100 bp
* 40465 45668: contig of 5204 bp in length
* 45669 45768: gap of 100 bp
* 45769 51359: contig of 5591 bp in length
* 51360 51459: gap of 100 bp
* 51460 56834: contig of 5375 bp in length
* 56835 56934: gap of 100 bp
* 56935 61504: contig of 4570 bp in length
* 61505 61604: gap of 100 bp
```



```

misc_feature      56935..61504
                  /note="assembly_fragment"
misc_feature      61605..67102
                  /note="assembly_fragment"
misc_feature      67203..72252
                  /note="assembly_fragment"
misc_feature      72353..78175
                  /note="assembly_fragment"

Query Match      8.0%; Score 22; DB 2; Length 149138;
Best Local Similarity 100.0%; Pred. No. 0.21;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      253 ATTTCGGAGCGCCGAGCGACGA 274
      |||||||
Db 144176 ATTTCGGAGCGCCGAGCGACGA 144197

RESULT 14
AC011492/c      154312 bp      DNA      linear      HTG 20-Jul-2001
LOCUS      Homo sapiens chromosome 19 clone CTB-187L3, *** SEQUENCING IN
DEFINITION      AC011492
ACCESSION      AC011492
VERSION      AC011492.7 GI:14971177
KEYWORDS      HTG; HTGS_PHASE2; HTGS_ACTIVEFIN.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 154312)
AUTHORS      DOE Joint Genome Institute and Stanford Human Genome Center.
JOURNAL      Direct Submission
TITLE      Unpublished
REFERENCE      2 (bases 1 to 154312)
AUTHORS      DOE Joint Genome Institute.
JOURNAL      Direct Submission
TITLE      Direct Submission
COMMENT
* Submitted (07-Oct-1999) Production Sequencing Facility, DOE Joint
  Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
  On Jul 20, 2001 this sequence version replaced gi:13752682.
* NOTE: This is a 'working draft' sequence. It currently
  * consists of 2 contigs. Gaps between the contigs
  * are represented as runs of N. The order of the pieces
  * is believed to be correct as given, however the sizes
  * of the gaps between them are based on estimates that have
  * been provided by the submittor.
  * This sequence will be replaced
  * by the finished sequence as soon as it is available and
  * the accession number will be preserved.
  *
  * 6785: contig of 6785 bp in length
  * 6786 6885: gap of unknown length
  * 6886 154312: contig of 147427 bp in length.
  -----Genome Center
          Center: Joint Genome Institute
          Center Code: JGI
          Web site: http://www.jgi.doe.gov
          -----
          Project Information
          Center Project Name: 136279
          Center clone name: CTB-187L3
          -----
          Summary Statistics
          Consensus quality: 152933 bases at least Q40
          Consensus quality: 152995 bases at least Q30
          Consensus quality: 153019 bases at least Q20
          Estimated insert size: 156750; agarose-fp estimation
          Estimated insert size: 153075; sum-of-contigs

          estimation
          Quality coverage: 14.63 in Q20 bases; agarose-fp
          estimation
          Quality coverage: 14.98 in Q20 bases; sum-of-contigs
          estimation
          Location/Qualifiers

```

source 1. .154312
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/clone="CTB-187L3"

BASE COUNT 35734 a 38388 c 38079 g 42009 t 102 others

ORIGIN

Query Match 8.0%; Score 22; DB 2; Length 154312;
Best Local Similarity 100.0%; Pred. No. 0.21;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 253 ATTTGCGAGCCGAGCAGCA 274
|||||
DB 38413 ATTTGCGAGCCGAGCAGCA 38392

RESULT 15
AC005609/c 157970 bp DNA linear PRI 04-SEP-1998
LOCUS Homo sapiens chromosome 5, BAC clone 203013 (LBNL.H155), complete
DEFINITION
ACCESSION AC005609
VERSION AC005609.1 GI:3540156
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 157970)
AUTHORS Kimerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
Kadner,K., Miguel,T., Miller,C., Pfluck,S., Pollard,M.,
Rojeski,H., Subramanian,S. and Martin,C.H.
TITLE Sequencing of human chromosome 5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 157970)
AUTHORS Ricker,D.O.
TITLE Large Scale Sequence Analysis and Annotation with the Sequence
Comparison Analysis (SCAN) System
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 157970)
AUTHORS Kimerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
Kadner,K., Miguel,T., Miller,C., Pfluck,S., Pollard,M.,
Rojeski,H., Subramanian,S. and Martin,C.H.
TITLE Direct Submission
JOURNAL Submitted (04-SEP-1998) Human Genome Center, DOE Joint Genome
Institute, Lawrence Berkeley National Laboratory, MS 74-157,
Berkeley, CA 94720, U.S.A.
COMMENT Sequence submitted by:
DOE Joint Genome Institute.
FEATURES
source Location/Qualifiers
1. .157970
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/map="5q"
/clone="BAC 203013"
/note="LBNL.H155"
/rpt_family="MLT"
complement(80. .214)
repeat_region /rpt_family="MLT"
complement(279. .410)
repeat_region /rpt_family="MSTc"
572. .835
repeat_region /rpt_family="L1"
1847. .2064
repeat_region /rpt_family="MER63A"
complement(3222. .3307)
repeat_region /rpt_family="Alu"
complement(3610. .3795)
repeat_region /rpt_family="MER20"
4687. .4996
repeat_region /rpt_family="Alu"
5251. .5270
repeat_region

/note="(r)20"
/rpt_type=tandem
/rpt_unit=r
complement(5254. .5543)
repeat_region /rpt_family="Alu"
5719. .5799
repeat_region /rpt_family="L1"
complement(7406. .9829)
CDS /codon_start=1
/product="KIAA0345-1like 1"
/protein_id="AAC34314.1"
/db_xref="GI:3540158"
/translation="MISWQGGPRPRQLLMLLILAAETGSGQLHYSEBAKRGTE
VGRIAODLGLLELAEVPLRFVRSKRHGDLLEVNILQNGILFVNSRIDREELGRASAC
SILHEIVDRPLQVPHVEYKVDINDNPDIPEESKRILIAESRPETRFCDGASDA
DIGVNSALTVELRDNDYFTLDAQNSLOEYKSLVLRKIDREELIHEHLLTASDGG
KPELGTVOLITLIDVNDNAPEFQSVKVLVLEAFNGTLYRLNATDPDQNGND
IYSSRRPWPVAVYAFITNPNGEIRTKGKIDPEEKKEIETSVAYDKGNTPMAGHC
TLLEVLVDNDNAPEYTTISLPTREDTQPSAITLISVSDSGSNQVTCITLPH
VPEKLVSTYKKNYSLVLSALDRESAYELVVTARDDGSPSLMTATSVSGVADVND
NAPARAOPEYTVFVEKNPQCHIFTVSAQDADQENALVSYSLVERRGEALSSYV
SVHASEGKYVALQPIDHELELLOFQVSARDGVPPLGSLNVLQVLENDNALLL
TPGASAGTVESELMPPRSYGAGHVAVKAVADAGVYKMWLSYELQAAVGRITFPR
GLYGEISTRTPLDEVDAPHRLVLVYKDHCPEALFATATVILSYESQARQASRA
SKGAVPEALVDVNYTLITACAVSLSLVLTLLTALRCAPPTBEGCARGKPTLV
CSAAGSWSYSQRRPRVCSGGEPKRTDLMARSPSLPCLGSAEETGOREDESECKE
VSLYFRKLS"
12033. .12446
repeat_region /rpt_family="MLT1"
13167. .13252
repeat_region /rpt_family="Alu"
13887. .13928
repeat_region /rpt_family="MER42"
complement(14247. .16625)
CDS /codon_start=1
/product="KIAA0345-1like 2"
/protein_id="AAC34315.1"
/db_xref="GI:3540159"
/translation="MVIIGPRGSGSORLLSLLLAAMEVSGQLHYSEBAKRGTE
VGRIAODLGLLELAEVPLRFVRSKRHGDLLEVNILQNGILFVNSRIDREELGRASAC
SILHEIVDRPLQVPHVEYKVDINDNPDIPEESKRILIAESRPETRFCDGASDA
DIGVNSALTVELRDNDYFTLDAQNSLOEYKSLVLRKIDREELIHEHLLTASDGG
KPELGTVOLITLIDVNDNAPEFQSVKVLVLEAFNGTLYRLNATDPDQNGND
IYSSRRPWPVAVYAFITNPNGEIRTKGKIDPEEKKEIETSVAYDKGNTPMAGHC
TLLEVLVDNDNAPEYTTISLPTREDTQPSAITLISVSDSGSNQVTCITLPH
VPEKLVSTYKKNYSLVLSALDRESAYELVVTARDDGSPSLMTATSVSGVADVND
NAPARAOPEYTVFVEKNPQCHIFTVSAQDADQENALVSYSLVERRGEALSSYV
SVHASEGKYVALQPIDHELELLOFQVSARDGVPPLGSLNVLQVLENDNALLL
TPGASAGTVESELMPPRSYGAGHVAVKAVADAGVYKMWLSYELQAAVGRITFPR
GLYGEISTRTPLDEVDAPHRLVLVYKDHCPEALFATATVILSYESQARQASRA
SKGAVPEALVDVNYTLITACAVSLSLVLTLLTALRCAPPTBEGCARGKPTLV
CSAAGSWSYSQRRPRVCSGGEPKRTDLMARSPSLPCLGSAEETGOREDESECKE
VSLYFRKLS"
18159. .18424
repeat_region /rpt_family="Alu"
complement(19526. .19773)
repeat_region /rpt_family="Alu"
complement(20562. .22994)
CDS /codon_start=1
/product="KIAA0345-1like 3"
/protein_id="AAC34316.1"
/db_xref="GI:3540160"
/translation="MFGFQRLGRTPRQDLMLLLEFWEVSGQLHYSEBAKRGTE
VGRIAODLGLLELAEVPLRFVRSKRHGDLLEVNILQNGILFVNSRIDREELGRASAC
SILHEIVDRPLQVPHVEYKVDINDNPDIPEESKRILIAESRPETRFCDGASDA
DIGVNSALTVELRDNDYFTLDAQNSLOEYKSLVLRKIDREELIHEHLLTASDGG
KPELGTVOLITLIDVNDNAPEFQSVKVLVLEAFNGTLYRLNATDPDQNGND
IYSSRRPWPVAVYAFITNPNGEIRTKGKIDPEEKKEIETSVAYDKGNTPMAGHC
TLLEVLVDNDNAPEYTTISLPTREDTQPSAITLISVSDSGSNQVTCITLPH
VPEKLVSTYKKNYSLVLSALDRESAYELVVTARDDGSPSLMTATSVSGVADVND
NAPARAOPEYTVFVEKNPQCHIFTVSAQDADQENALVSYSLVERRGEALSSYV
SVHASEGKYVALQPIDHELELLOFQVSARDGVPPLGSLNVLQVLENDNALLL
TPGASAGTVESELMPPRSYGAGHVAVKAVADAGVYKMWLSYELQAAVGRITFPR
GLYGEISTRTPLDEVDAPHRLVLVYKDHCPEALFATATVILSYESQARQASRA
SKGAVPEALVDVNYTLITACAVSLSLVLTLLTALRCAPPTBEGCARGKPTLV
CSAAGSWSYSQRRPRVCSGGEPKRTDLMARSPSLPCLGSAEETGOREDESECKE
VSLYFRKLS"

```
AGASDEPALVDVNVYLLIAICVSSLLVTLTLTALMMSATPTGACGKPTLVC
SRAVGSMSYQQRORVCSSEGPCKTDLMAFSPLGLNKEEGEQRQESNMPCGV
SFQIDPPIRCKM"
repeat_region complement(27063, .27162)
/rpt_family="Alu"
repeat_region complement(28088, .28385)
/rpt_family="Alu"
repeat_region complement(31070, .31371)
/rpt_family="Alu"
repeat_region complement(31704, .31779)
/rpt_family="Alu"
repeat_region complement(32012, .32081)
/rpt_family="Alu"
CDS complement(33515, .36049)
/codon_start=1
/product="KIAA0345-1like 4"
/protein_id="AAC34317.1"
/db_xref="GI:3540161"
/translation="MVSRCSCIGVQCCLLSLLLAAMEVSGQLHYSVPEARHGTFV
GRIADLGLLELAEVORLFRVASKRHGDLEVNLONGILFVNSRIDREELCGRSVCS
IHELVIVDRPIQVPHVDVEVKDINDNPVPAOKMLFTAESRPLDRPPLGASDA
VGENALITTKLSNREYFVLDIINKDKDKFPVLYLRLDRENPQLKLLIATDGGK
PEFTGVSLLIVLDANDNP IFRPYEYVKMYENQVOTVLRNLASDSDEGINKEK
MYSESSLVPTIRRKFWINERTGEIKVNDADIDEDSNTYEIHVDYDGNPMPYHCT
VLVELLDENDNSPEVITYSLSPYKEDAQGYVIALISVDHDSGANGOVTCSLPHV
PFKLYSTYKNYSVLVDSALDRERVSAYELVYARODGSPIMATASVSEVADVND
APAAQSEVYFVYKNNPQCHIFTVSAMDADAENALVYSILVERLGERLSSTYS
VHAESGKYALQPLDHELELDFOVSARDGVPPILGSNLTLOVFLVENDNAPALA
SPAGSAGVASELVLSVAVGHVAVKRAVDADSGYNAMLSTELQSAVAGARIPEFRG
LYTGEISTRALDETDSPQRLVLVKGDEPSTLTATVLSLVESGQAPKASSRAS
VGVAPEVALVDVNVYLLIAICAVSSLLVTLTLTLTLALRCSAAPTEGACGPKPLVCS
SAVGSMSYQQRORVCSSEGLPKADMAFSPSLPCPMYDVGDEQDSIGDHSRAKVG
YYVFIFLCCFMNNIFSYRIFSNMYONISFLSTFHLCLNISDPFTVI"
repeat_region complement(37922, .38185)
/rpt_family="Alu"
gene complement(37939, .44326)
/sequence="KIAA0345-1like 5"
38925, .39212
repeat_region complement(41074, .43602)
/rpt_family="Alu"
CDS complement(41074, .43602)
/sequence="KIAA0345-1like 5"
/codon_start=1
/product="KIAA0345-1like 5"
/protein_id="AAC34313.1"
/db_xref="GI:3540157"
/translation="MLYSSRGDEPGQPLLSTLLIAMVVGSGQLHYSVPEAEHGT
VGRADLGLLELAEVRLFDLSDSGDLEVNLONGILFVNSRIDREELCGRSAC
SILHEVIVDRPIQVPHVDVEVKDINDNPVPAOKMLFTAESRPLDRPPLGASDA
DIGENALITTKLSNREYFVLDIINKDKDKFPVLYLRLDRENPQLKLLIATDGG
KPELTGVLQTLITVLDNDNAPVDFRTLTVKLPENYIGTLVTHPNASDLDEGLND
IITYSSSDVSPDIKSKFMDPLSGAIVIGHMDFEESRAKIPYEAVDKGFPLGHC
TLLVEVDVNDNAPOLITIKTSLPVKEDAGVIALISVIDLADANGOVTCSLPH
VPEKLYSTYKNYSVLVDRALDRESVAYELVYARODGSPSLMATASVSEVADVND
NAPAAQSEVYFVYKNNPQCHIFTVSARDADAENALVYSILVERLGERLSSTYS
SVHAESGKYALQPLDHELELDFOVSARDGVPPILGSNLTLOVFLVENDNAPALA
TPMRGTGAVSEVLRSGVAGVYVGVKRAVDADSGYNAMLSTELQETASATPEFRV
GLYTGEISTRALDETDAPQRLVLVKGDEPALTATVLSLVESGQAPKSSRA
SVGATGPEVTLVDVNVYLLIAICAVSSLLVTLTLTLTLALRCSAAPTEGACGPKPLV
CSSAVGSMSYQQRORVCSSEGEKOKTDLMAFSPGLSPGASSTERTGEPSSASDSTGK
VGFSSILFIYIIFLEERYRLPLGAVQIVLFLEIQIIFLIK"
45429, .45723
repeat_region complement(48332, .50776)
/rpt_family="Alu"
CDS complement(48332, .50776)
/codon_start=1
/product="KIAA0345-1like 6"
/protein_id="AAC34318.1"
/db_xref="GI:3540162"
/translation="MDYHMRGELSGRRLLLLLLLLAAMKVGSGQLHYSVPEAKHGT
VGRADLGLLELAEVRLFRVASKRHLDLEVSLONGILFVNSRIDREELCGRSAC
SILHEVIVDRPIQVPHVDVEVKDINDNPVPAOKMLFTAESRPLDRPPLGASDA
DVGANSVLTFRLLSHDYFMDLVNSKNDENKLVLRKSLDRBDAPAHHLFLATDGG
```

Query Match 8.0%; Score 22; DB 9; Length 157970;
Best Local Similarity 100.0%; Pred. No. 0 21;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 253 ATTTCGGAGGCGCGAGCAGGA 274
|||||
Db 53409 ATTTCGGAGGCGCGAGCAGGA 53388

Search completed: October 8, 2002, 14:01:52
Job time : 2377 secs

